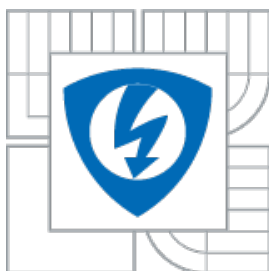




**VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ**  
BRNO UNIVERSITY OF TECHNOLOGY



**FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH  
TECHNOLOGIÍ**  
**ÚSTAV BIOMEDICÍNSKÉHO INŽENÝRSTVÍ**

**FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION**  
**DEPARTMENT OF BIOMEDICAL ENGINEERING**

# **MATEMATICKÁ ANALÝZA DAT V HMOTNOSTNÍ SPEKTROMETRII**

**MATHEMATICAL ANALYSIS IN MASS SPECTROMETRY**

**BAKALÁŘSKÁ PRÁCE**  
BACHELOR'S THESIS

**AUTOR PRÁCE**  
AUTHOR

**ELIŠKA HUTOVÁ**

**VEDOUCÍ PRÁCE**  
SUPERVISOR

**ING. MARTIN VALLA**

**BRNO 2010**



**VYSOKÉ UČENÍ  
TECHNICKÉ V BRNĚ**

**Fakulta elektrotechniky  
a komunikačních technologií**

**Ústav biomedicínského inženýrství**

# **Bakalářská práce**

bakalářský studijní obor

**Biomedicínská technika a bioinformatika**

**Studentka:** Eliška Hutová

**ID:** 106151

**Ročník:** 3

**Akademický rok:** 2009/2010

## **NÁZEV TÉMATU:**

**Matematická analýza dat v hmotnostní spektrometrii**

## **POKYNY PRO VYPRACOVÁNÍ:**

Seznamte se s principy analýzy látek pomocí hmotnostní spektrometrie a tandemové hmotnostní spektrometrie. Seznamte se s volně dostupnými nástroji pro analýzu hmotnostních spekter získaných pomocí obou metod. Analyzujte vzorek flavonoidní sloučeniny pomocí hmotnostního spektrometru a spektrum popište. Prostudujte modelová spektra vybraných flavonoidních sloučenin. Identifikujte naměřený vzorek pomocí matematických nástrojů. Výsledky vhodnou formou interpretujte.

## **DOPORUČENÁ LITERATURA:**

- [1] Kinter, M., Sherman, N.: Protein Sequencing and Identification Using Tandem Mass Spectrometry . 1st edition. New York : Wiley, John & Sons, Incorporated, 2000. 301 s. New Edition. ISBN 9780471322498.
- [2] Kapp, E., Schutz, F.: Overview of Tandem Mass Spectrometry (MS/MS) Database Search Algorithms. Protein Sci.: UNIT 25.2 [online]. 2007 [cit. 2009-05-15].

**Termín zadání:** 8.2.2010

**Termín odevzdání:** 31.5.2010

**Vedoucí práce:** Ing. Martin Valla

**prof. Ing. Ivo Provazník, Ph.D.**

*Předseda oborové rady*

## **UPOZORNĚNÍ:**

Autor bakalářské práce nesmí při vytváření bakalářské práce porušit autorská práva třetích osob, zejména nesmí zasahovat nedovoleným způsobem do cizích autorských práv osobnostních a musí si být plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č.40/2009 Sb.

## Abstrakt

V této bakalářské práci je teoreticky popsán princip metody hmotnostní spektrometrie a tandemové hmotnostní spektrometrie, využití metody v praxi a popis funkčnosti jednotlivých analyzátorů. Dále je uveden popis hmotnostního spektra a algoritmy pro jeho analýzu. Praktická část práce popisuje tvorbu analytického nástroje pro vyhodnocování spekter a následné použití tohoto programového nástroje pro praktickou analýzu flavonoidních sloučenin, konkrétně isoflavonů.

## Abstract

In this bachelor work, there is theoretically described the principle of mass spectrometry and tandem mass spectrometry, usage the method in practices and description of functionality of separate analysers. Next, there is also mentioned the description of mass spectrum and the algorithm to its analysis. Practical part of the work describes the creation of analytic tool to evaluate spectrum and following usage of this programming tool for the practical analysis of flavonoid compounds, concretely isoflavonoid.

## Klíčová slova

Hmotnostní spektrometrie, tandemová hmotnostní spektrometrie, MS, MS/MS, hmotnostní spektrometr, tandemový hmotnostní spektrometr, hmotnostní spektrum, flavonoidní sloučeniny, isoflavony, t2d, Matlab, Java, Proteom

## Keywords

Mass spectrometry, tandem mass spectrometry, MS, MS/MS, mass spectrometr, tandem mass spectrometr, mass spectrum, flavonoid compounds, isoflavons, t2d, Matlab, Java, Proteom

## Citace

HUTOVÁ, E. *Matematická analýza v hmotnostní spektrometrii*. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, 2010. 53 s. Vedoucí bakalářské práce Ing. Martin Valla.

## **Prohlášení**

Prohlašuji, že svoji bakalářskou práci na téma Matematická analýza dat v hmotnostní spektrometrii jsem vypracovala samostatně pod vedením vedoucího bakalářské práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou všechny citovány v práci a uvedeny v seznamu literatury na konci práce.

Jako autorka uvedené bakalářské práce dále prohlašuji, že v souvislosti s vytvořením této práce jsem neporušila autorská práva třetích osob, zejména jsem nezasáhla nedovoleným způsobem do cizích autorských práv osobnostních a jsem si plně vědoma následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., včetně možných trestněprávních důsledků vyplývajících z ustanovení § 152 trestního zákona č. 140/1961 Sb.

V Brně dne 31. května 2010

.....  
podpis autora

## **Poděkování**

Děkuji vedoucímu bakalářské práce Ing. Martinu Vallovi za jeho trpělivost a toleranci a také za přínosnou metodickou, pedagogickou a odbornou pomoc, za cenné rady a náměty při zpracování mé bakalářské práce. Dále bych chtěla poděkovat doc. Ing. René Kizkovi, Ph.D., z Mendlovy univerzity v Brně za jeho odborné postřehy, ochotu při realizaci bakalářské práce a také za pomoc při získávání zpracovávaných dat.

V Brně dne 31. května 2010

.....  
podpis autora

## Obsah

1	Úvod .....	4
2	Hmotnostní spektrometrie .....	5
2.1	Princip metody hmotnostní spektrometrie.....	5
2.2	Hmotnostní spektrometr .....	6
2.3	Ionizace .....	6
2.4	ESI.....	7
2.5	MALDI.....	8
2.6	Hmotnostní analyzátory.....	8
2.6.1	Základní funkce hmotnostního analyzátoru: .....	8
2.6.2	Základní typy hmotnostních analyzátorů: .....	8
2.7	Příprava vzorku pro MS .....	9
2.7.1	Příprava vzorku pro metodu MALDI .....	9
2.7.2	Příprava vzorku pro metodu ESI .....	9
2.8	Výsledky MS .....	10
2.8.1	Hmotnostní spektrum .....	10
2.8.2	Přesnost určení hmoty .....	10
2.8.3	Definice jednotky Dalton .....	11
2.9	Možnosti identifikace hmotnostních spekter.....	11
2.9.1	Databázové vyhledávací programy .....	11
3	Tandemová hmotnostní spektrometrie .....	12
3.1	Princip metody TMS .....	12
3.2	Tandemový hmotnostní spektrometr.....	12
3.3	Fragmenty spekter .....	13
3.4	Výsledky z MS/MS .....	15
3.4.1	Rozdíl mezi výsledky z MS a MS/MS .....	15
3.5	Přehled možností identifikace MS/MS dat.....	16
3.5.1	„De Novo“ peptidové sekvenování .....	16
3.5.2	Identifikace s využitím tagů .....	17
3.5.3	Metoda fragmentového mapování.....	17
3.6	X!Tandem.....	17
3.6.1	Pracovní postup .....	18
3.6.2	Schéma skórování algoritmu X!Tandem.....	19

3.6.3	Histogram hyper-skóre .....	20
3.6.4	Výpočet pravděpodobnostního koeficientu shody (E-value) .....	21
3.6.5	Binomické rozložení.....	23
3.7	OMSSA .....	24
3.7.1	Pracovní tok algoritmu OMSSA .....	24
3.7.2	Výpočet E-value .....	28
3.7.3	Reskórování ke zlepšení citlivosti .....	29
4	Datové struktury .....	30
4.1	mzML formát ve vztahu k MS/MS .....	30
4.2	mzXML formátu ve vztahu k MS/MS.....	30
4.3	t2d formát ve vztahu k MS .....	30
4.4	Proteom .....	30
5	Flavonoidní sloučeniny .....	31
5.1	Chemická stavba flavonoidů .....	31
5.2	Isoflavonoidy.....	31
5.2.1	Dělení isoflavonoidů do podskupin.....	32
5.2.2	Isoflavony.....	32
5.2.3	Výskyt isoflavonů.....	33
5.2.4	Isoflavony v rostlinách čeledi bobovité.....	33
5.2.5	Isoflavony v rostlinách čeledi konopovité.....	36
5.3	Izolace a identifikace isoflavonů .....	37
5.4	Sladina a mladina ve výrobě piva.....	38
6	Realizace aplikace .....	39
6.1	Požadavky na aplikaci pro analýzu .....	39
6.2	Hmotnostní spektrometrie v bioinformatickém toolboxu programu Matlab.....	39
6.3	Popis implementace.....	40
6.3.1	Způsob načtení t2d dat.....	40
6.3.2	Zobrazení teplotních map (heat map) a rozpoznání píků (peak recognize).....	41
6.3.3	Získání sloučených spekter.....	42
6.3.4	Uložení získaných dat do souboru csv .....	42
6.3.5	Uložení získaných statistických parametrů .....	42
7	Popis aplikace a zpracování výsledků .....	44
7.1	Spektra testovacích a modelových isoflavonů.....	44

7.2	Statistické parametry v analýze spekter .....	44
7.2.1	Rozptyl $D(x)$ .....	44
7.2.2	Střední hodnota $E(x)$ .....	44
7.2.3	Korelační koeficient $\rho_{xy}$ .....	44
7.3	Vyhodnocení metody porovnání spekter metodou aritmetického průměru .....	45
8	Závěr .....	48
	Seznam literatury .....	49
	Seznam obrázků .....	52
	Seznam tabulek .....	53
	Seznam příloh .....	53

# 1 Úvod

Biomedicínské inženýrství a bioinformatika se v poslední době staly nedílnou součástí medicínského, biologického i technického výzkumu. Spojením těchto interdisciplinárních oborů dochází k velkému rozvoji především na poli lékařském a přírodovědeckém, kdy mohou být hmotnostní spektrometry klasické i tandemové využity k screeningu nemocí, zkoumání biologických látek a to jak živočišného, tak i rostlinného původu.

Nejdůležitějším a v dnešní době nejvíce rozvíjejícím se oborem využití je však bioinformatika v kombinaci s genetickým inženýrstvím a výzkumem, kde jsou hmotnostní spektrometry a tandemové hmotnostní spektrometry využívány k mapování genomu a to nejen lidského.

Hmotnostní spektrometr je přístroj, který pomocí poměrně jednoduché detekční metody separuje urychlené částice ve vakuu podle jejich hmotnosti při průchodu magnetickými a elektrickými poli. Tandemový hmotnostní spektrometr je využíván především k zisku proteomických dat, kdy jeho princip spočívá zejména v porovnání získané databáze s databází vzorovou. Hodnocení získaných proteomů může být provedeno několika způsoby. V práci jsem se blíže seznámila se dvěma těmito algoritmy, takzvaně *open-source*, X!Tandem a OMSSA.

Flavonoidní sloučeniny jsou látky rostlinného původu. Dělí se podle své chemické stavby do několika skupin, z nichž pro člověka nejpřínosnější, a proto i nejvýznamnější jsou isoflavony. O působení těchto látek na lidské tělo bylo napsáno již mnoho odborných publikací a jejich pozitivní vliv je tedy více než dobře znám.

Ve své práci jsem měla za úkol vytvořit program, který dokáže spektra získaná v hmotnostním spektrometru z isoflavonů analyzovat a statisticky vyhodnotit. Tato analýza byla provedena v programu Matlab a následné porovnání v MS Excel.



## 2 Hmotnostní spektrometrie

Hmotnostní spektrometrie (MS) je fyzikálně-chemická detekční metoda, která využívá separace urychlených ionizovaných částic (iontů) ve vakuu podle jejich hmotnosti při jejich průchodu magnetickými a elektrickými poli.

Metoda byla vyvinuta počátkem 20. století a původně byla využívána zvláště ve fyzice a chemii, mj. byly pomocí této metody objeveny stabilní izotopy prvků.

Během posledních dvaceti let se použití metody rozšířilo snad do všech oblastí přírodních věd, od geologie, přes fyziku, astronomii a chemii k biologii, k analýze životního prostředí, do farmacie, medicíny. Nyní se s hmotnostní spektrometrií můžeme setkat buď přímo, např. během osobních bezpečnostních prohlídek na letištích, nebo alespoň nechtěně zprostředkovaně, kdy tato metoda je použita např. v lékařství (diagnostika nádorových onemocnění, metabolických chorob atd.). Metoda je nezastupitelná pro kriminalisty (identifikace drog, detekce výbušnin, identifikace pachatelů zvláště nebezpečné trestné činnosti), je výhradním nástrojem antidopingových kontrol, používá se v archeologii (datování stáří předmětů) a geologii, v monitorování kvality životního prostředí, v toxikologii, chemii (sledování průběhu chemických reakcí). Hmotnostní spektrometry jsou součástí výbavy většiny vesmírných sond i některé speciální vojenské techniky.

Za vývoj technik hmotnostní spektrometrie byly uděleny celkem 4 Nobelovy ceny za fyziku nebo chemii (1906, 1922, 1990, 2002) a využití této techniky bylo zásadní pro udělení nejméně dvou dalších Nobelových cen [1].

Hmotnostní spektrometrie je součástí studia proteomu.



**Obrázek 1:** Hmotnostní spektrometr, THERMO FISHER LTQ Orbitrap XL [2]

### 2.1 Princip metody hmotnostní spektrometrie

Princip metody spočívá v rozdělení nabitých částic podle jejich molekulových hmotností. Hmotnostní spektrometr separuje nabitě částice podle jejich měrného náboje  $m/z$  ( $m$  je hmotnost,  $z$  je náboj) a umožňuje jejich stanovení. Dále poskytuje údaje o relativním zastoupení stejného  $m/z$  ve směsi iontů. Záznam molekulárních a fragmentovaných iontů je charakteristický pro danou látku (protein) a dává cenné informace o její struktuře a na jeho základě lze většinou strukturu látky odvodit nebo potvrdit. Hmotnostní spektrometrie je citlivá metoda a umožňuje analyzovat látky v množství až

$10^{-15}$  g [5]. Všechny tyto operace probíhají v uzavřeném prostoru, kde je kontinuálně udržováno vakuum.

K identifikaci proteinu lze v zásadě využít dva základní způsoby: jednoduchá a tandemová hmotnostní spektrometrie. U jednoduché MS je protein naštěpen proteolytickým enzymem na menší peptidy, jejichž přesné hmotnosti jsou pomocí MS změřeny. Spektrum těchto hmotností je pak porovnáváno s teoretickými spektry, která jsou vypočítána ze sekvencí proteinů v dostupných databázích. Tandemová MS využívá dvou jednoduchých hmotnostních spektrometrů spolupracujících v spolupracujících v kooperaci (tandemu) [4].

## 2.2 Hmotnostní spektrometr

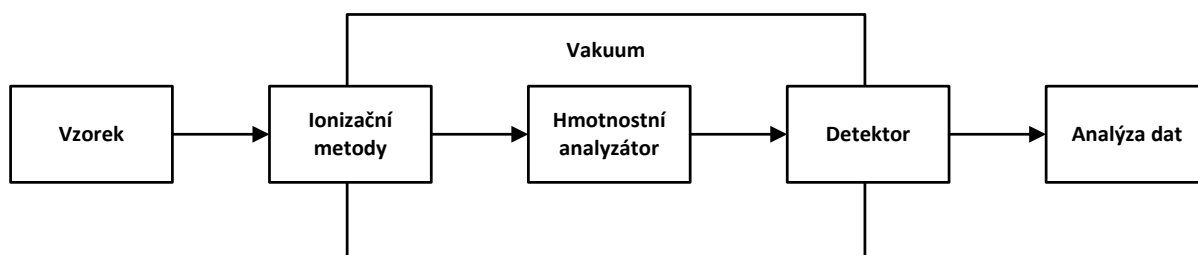
Metoda hmotnostní spektrometrie se realizuje na přístroji zvaný hmotnostní spektrometr.

Hmotnostní spektrometr je iontově – optické zařízení, jehož základní obecné uspořádání složeno ze tří hlavních částí (Obrázek 2):

- Iontový zdroj – štěpí molekuly na ionty. (Elektronová ionizace – „tvrdá“ technika, měkké ionizační techniky, viz níže)
- Hmotnostní analyzátor – separuje ionty podle poměru  $m/z$  při aplikaci elektromagnetického pole. Řídí se Newtonovými zákony a Lorenzovou silou. (viz níže)
- Detektor – poskytuje data pro výpočet množství každého iontu, který na něj dopadne. (Druhy detektorů: elektronový, fotonásobič, Faradayova klec).

Hmotnostní analyzátor musí vždy pracovat za vysoké hodnoty vakua, ta se liší podle typu analyzátoru v rozmezí  $10^{-3}$  až  $10^{-10}$  Pa. Takto vysoká hodnota vakua je nutná z důvodu toho, že ionty musí mít dostatečně dlouhou střední dráhu a nesmí docházet ke kolizním srážkám s neutrálními atomy [3].

Hmotnostních spektrometrů v současnosti existuje celá řada, jejich konkrétní typy často vznikají kombinacemi specifických zařízení použitých pro iontový zdroj a hmotnostní analyzátor.



**Obrázek 2:** Základní schéma hmotnostního spektrometru

## 2.3 Ionizace

Jelikož zkoumaná látka musí být převedena jako intaktní do plynné fáze, využití hmotnostní spektrometrie pro analýzu proteinů bylo umožněno vývojem "měkkých" ionizačních technik hmotnostní spektrometrie, kam se řadí optimální techniky *matrix laser desorption/ionization* (MALDI) a *elektrospray ionization* (ESI). Pro optimální zpracování rozličných druhů látek je potřeba další možnosti ionizace. Další možné techniky ionizace (*jedná se především o techniky tvrdé ionizace*)[6]:

- Náraz elektronů (EI) – tato technika se dnes již nedoporučuje, protože zde ve skutečnosti nedochází k nárazu elektronu do molekuly, ale pouze k ovlivnění elektromagnetických polí.
- Působení elektrostatického pole (FI, FD).
- Chemická ionizace (CI).
- Nárazem rychlými atomy nebo ionty (FAB).
- Ionizace fotony.
- Ionizace  $^{252}\text{Cf}$ .

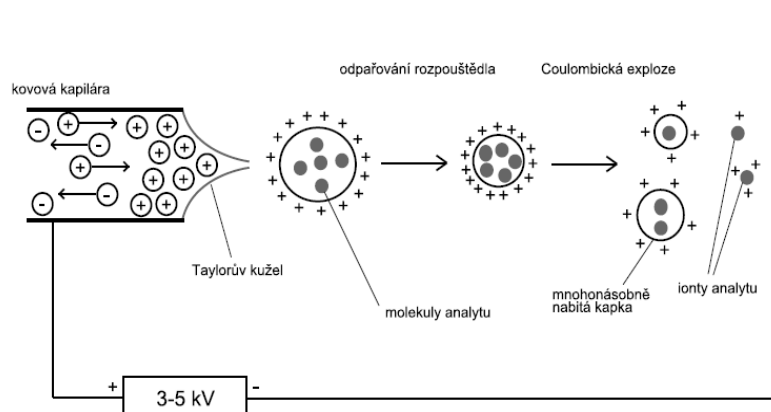
## 2.4 ESI

*Electrospray ionization* (ionizace elektrosprejem). Ionizační technika, která produkuje vícenásobně nabitě ionty z proteinů nebo peptidů. Metoda nevyžaduje předchozí fragmentaci (rozdělení zkoumané látky na menší částice). Ze záznamů ESI spektra lze tak získat přesné informace o molekulových hmotnostech analyzovaných látek. Při použití ESI techniky je možné zvýšit fragmentaci až na jednotlivé ionty. Typicky lze použít například s iontovou pastí [7].

Roztok vzorkuje přiváděn do iontového zdroje kapilárou, na kterou je přiváděno elektrické napětí řádově několika kV. Přívod napětí je umožněn prostřednictvím přídavné kapaliny, kapalinovým spojem s vloženou elektrodou nebo přes pokovený hrot kapiláry. Druhý pól napětí je přiváděn na elektrodu se vstupním otvorem. Vlivem elektrického pole vzniká sprej aerosolu pohybující se směrem ke vstupnímu otvoru a je tvořen nabitými kapičkami rozpouštědla s analytem. Za atmosférického tlaku, při použití protiproudů sušícího plynu nebo i bez něj, se rozpouštědlo odpařuje a přenosem náboje vznikají nabitě ionty analytu, vstupující přes štěrby diferenciálního pumpování (sampler, skimmer) do analyzátoru [8].

Pro ESI je typický vznik vícenásobně nabitých iontů; ve spektrech jsou pozorovány charakteristické „obálky“ píků pro jednotlivé analyty, odpovídající iontům stejné látky s různým nábojem. Vznikají převážně vícenásobně nabitě ionty  $[\text{ABC}+n\text{H}]^{n+}$ , dále  $[\text{ABC}+\text{Na}]^+$ ,  $[\text{ABC}+\text{K}]^+$  a v nízkém rozsahu i fragmenty. Přítomnost více signálů od téhož analytu umožňuje přesné určení molekulové hmotnosti částice, klesá však citlivost metody v důsledku celkově nižších intenzit signálů.

Ionizace ESI je měkká ionizační technika umožňující analýzu částic o hmotnosti až  $10^6$  Da a patří, spolu s MALDI, ke dvěma nejužívanějším způsobům ionizace pro analýzu biopolymerů [9].



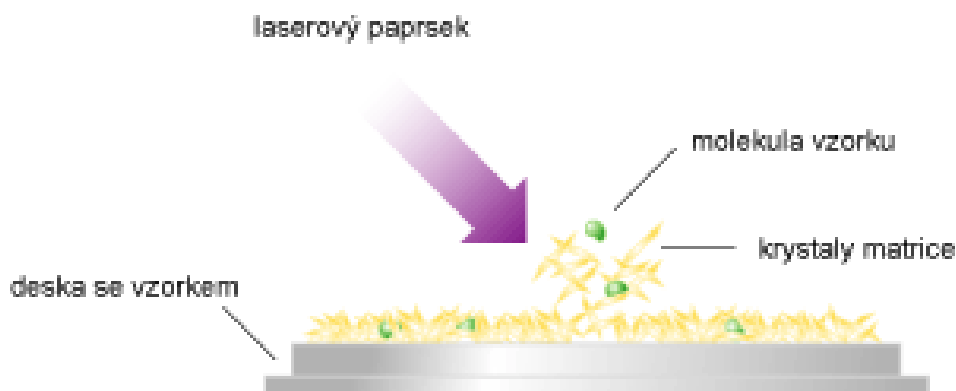
Obrázek 3: Princip metody ESI [10]

## 2.5 MALDI

*Matrix assisted laser desorption/ionization* (desorpce laserem za přítomnosti matrice). Tato metoda se používá pro analýzu biomolekul a velkých organických látek, které jsou náchylné k fragmentaci při ionizaci. Ke stanovení vyšších molekulových hmotností se používá společně kombinaci s detektorem doby letu TOF (*time-of-flight*). Detektor umožňuje změřit dobu průletu a z ní lze vypočítat rychlost částice. Ionty analyzované látky jsou urychleny silným elektrickým polem. Zařízení je tvořeno destičkou s terčíkem, do kterého se aplikuje vzorek a matrice způsobí jeho krystalizaci (Obrázek 4).

Směs matrice a vzorku v pevném stavu a na vhodném nosiči, například na nerezové destičce, je zasažena nanosekundovým pulzem laseru. Matrice energii pulzu absorbuje a její rozklad ionizuje molekuly vzorku. Touto ionizací se rozumí adice kationtu ( $H^+$ ,  $Na^+$ ) či aniontu na molekulu vzorku, disociace  $H^+$  z molekuly vzorku, vznik radikálu odštěpením elektronu, popřípadě cílené rozkouskování (vysokou energií laseru) molekuly vzorku a opět spojení kousků. Ionty analyzované látky jsou urychleny silným elektrickým polem (25–30 kV) a přes uzemněnou mřížku vstupují do vakua v trubici detektoru letu, kde se pohybují rychlostí danou jejich hmotností a nábojem. Zde se měří doba letu částice, z níž se pak vypočte poměr molekulové hmotnosti a náboje částice.

Hmotnostní spektroskopie MALDI byla původně vyvinuta pro kvalitativní analýzu peptidů a bílkovin, avšak nyní se využívá i pro analýzy nukleových kyselin nebo nízkomolekulárních organických i anorganických látek. Výhodou je vysoká citlivost a rychlost měření [11].



Obrázek 4: Schéma metody MALDI [12]

## 2.6 Hmotnostní analyzátoři

### 2.6.1 Základní funkce hmotnostního analyzátoru:

- Slouží k rozdělení iontů podle jejich poměru  $m/z$ , kvalitu rozdělení iontů vyjadřuje rozlišovací schopnost.
- Je umístěn za iontovým zdrojem (neutrální molekuly jsou už převedeny na ionty) a před detektorem (před detekcí je třeba rozdělit ionty podle  $m/z$ ).

### 2.6.2 Základní typy hmotnostních analyzátorů:

1. Magnetické analyzátoři - v magnetickém (nebo elektrickém) poli dochází k zakřivení dráhy letu iontů.

2. Kvadrupólové a iontové pasti – separace nastává podle různé stability oscilace iontů v dvoj- nebo trojrozměrné kombinaci vysokofrekvenčního střídavého napětí.
3. Analyzátoři doby letu TOF – urychlené ionty se v oblasti bez pole pohybují různou rychlostí v závislosti na hodnotě  $m/z$  (čím menší iont, tím rychlejší doba letu). Ionty jsou urychleny na vstupu do analyzátoru a poté je změřen čas, za který „dolétnou“ k detektoru, čímž je určena hodnota jejich poměru  $m/z$ .
4. Iontová cyklotronová rezonance ICR - separace nastává podle různé absorpce energie při cyklonálním (pohyb po kruhové dráze s frekvencí  $\omega$  nepřímo úměrnou  $m/z$ ) pohybu iontů v kombinovaném magnetickém a elektrickém poli.

## 2.7 Příprava vzorku pro MS

Měření molekulových hmotností molekul (přesněji jejich různě nabitých iontů) má vždy několik kroků [6]:

1. Převedení molekuly do plynné fáze (vysoké vakuum), tím molekula získá charakteristický náboj.
2. Urychlení iontu. Pomocí charakteru jeho pohybu uvnitř vakuového prostoru lze vypočítat poměr jeho hmotnosti a náboje.
3. Určení parametrů, které charakterizují dráhu iontu pomocí detektoru (viz bod 2).
4. Zpracování signálu vycházejícího z detektoru a vypočítání poměru hmotnosti a náboje příslušných iontů ( $m/z$ ) pomocí navazujícího elektronického systému.

### 2.7.1 Příprava vzorku pro metodu MALDI

Při této metodě je vzorek zkoumané látky ukotven na nerezové destičce v netěkavé matrici (kokrystalizace). Vhodná látka k použití je například kyselina nikotinová nebo kyselina 2,5 - dihydroxybenzoová. Vzorek (1 mg/ml) se nanese v množství 0.5 ml na nerezovou destičku a nechá se vysušit. Pak se aplikuje 0.5 ml matrice a opět se nechá vysušit. Matrice se volí dle vzorku, účelem použití je desorpce energie laseru. Destička se vloží do přístroje, zacílí se laserový paprsek („fire“), transfer energie způsobí ionizaci. Směřovaný energetický impuls poskytuje vysoké výtěžky iontů intaktního analytu a je dosaženo subpikomolární sensitivity. Dodáním energie dojde k odpaření matrice, která se nachází v nadbytku; v plynné fázi pak matrice nese analyt. Analyt je tak převeden do plynné fáze nepřímo. Matrice je zároveň donorem či akceptorem protonu, podle modu ionizace. Vůbec první byla kyselina nikotinová [13].

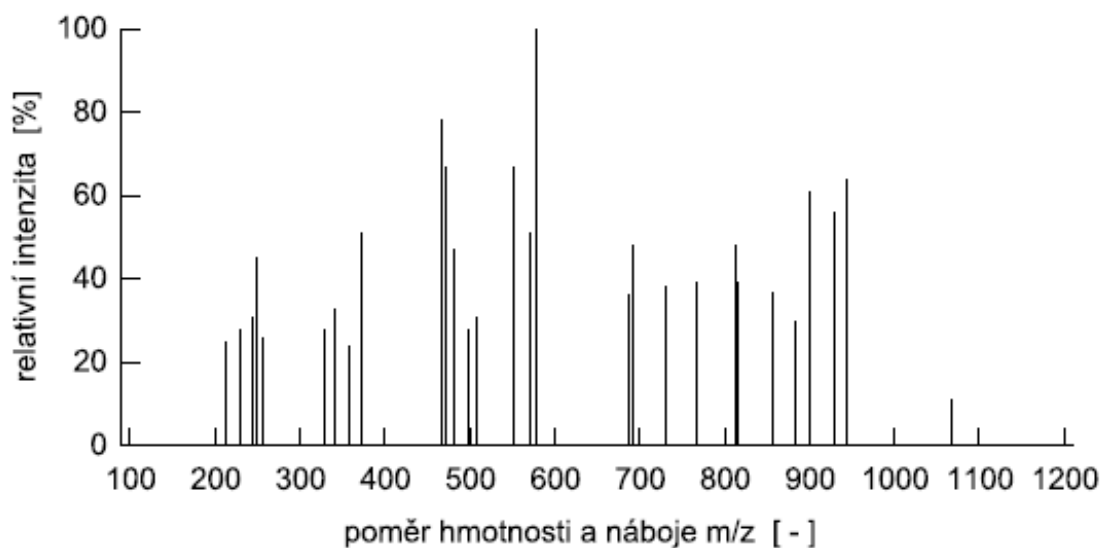
### 2.7.2 Příprava vzorku pro metodu ESI

Způsob provedení této metody spočívá v tom, že vzorek je rozpuštěn v těkavém rozpouštědle a rozprašován pomocí mikrostříkačky (kovová, skleněná s kovovým pístem). Vznikne tak aerosol drobných kapiček, který je vysoušen proudem suchého dusíku. Jak klesá velikost kapičky, tak roste hustota náboje. Dojde k tzv. kulombické explozi, při které se uvolní ionty, které odchází do spektrometru. Poté je vzorek disintegrován na jednotlivé molekuly a ionizován při velmi jemných podmínkách. V ideálním případě by vzorek měl být rozpuštěn v čistém rozpouštědle, u reálných vzorků biomakromolekul je někdy nutné zachovat pufrů kvůli stabilitě, ale musí být velmi zředěné. U ESI je velmi výhodné použití např. uhličitanu amonného. Ionty z jiných pufrů mohou ve spektru interferovat se studovanou látkou [13].

## 2.8 Výsledky MS

### 2.8.1 Hmotnostní spektrum

Hmotnostní spektrum zobrazuje závislost četnosti výskytu jednotlivých typů iontů na poměru jejich hmotnosti a náboje  $m/z$ . Protože hodnoty absolutní intenzity iontů jsou často i řádově závislé na konkrétním způsobu měření, převádí se četnost do normalizovaného tvaru (na relativní intenzitu vyjádřenou v %). Jednotlivé typy iontů s daným poměrem  $m/z$  jsou v grafickém zobrazení spektra viditelné jako „*peaky*“ (píky). Nejvyššímu píku ve spektru pak přísluší hodnota intenzity 100% (Obrázek 5).



Obrázek 5: Hmotnostní spektrum [14]

Současné metody identifikace proteinů využívají obvykle hmotnostní spektra získaná „jednoduchou“ MS nebo tandemovou MS/MS hmotnostní analýzou. V případě hmotnostní spektrometrie jsou „nastříhané“ části proteinu (peptidy) ionizovány, čímž se z jejich neutrálních molekul stávají ionty, které jsou následně zaznamenány ve výsledném spektru.

Z principu hmotnostní spektrometrie vyplývá, že neionizované peptidy se nemohou dostat k detektoru a tudíž nejsou zaznamenány.

Tandemová hmotnostní spektrometrie jde oproti MS analýze ještě o krok dále a pro každý typ iontu vygeneruje hmotnostní spektrum jeho fragmentů. Při použití této metody tedy nezískáme jedno spektrum ionizovaných peptidů jako u MS analýzy, ale sadu fragmentových spekter pro jednotlivé typy iontů [14].

### 2.8.2 Přesnost určení hmoty

Parametr hmotnostního analyzátoru

1. **Absolutní** – udává se v Daltonech (Da), hodnoty 0.1 – 0.0001
2. **Relativní** (mění se podle  $m/z$ ) – udává se v % nebo ppm (*parts per million*), hodnoty 100 – 0.1 ppm

Vyjadřuje shodu mezi naměřenou  $m/z_{\text{měřená}}$  a vypočtenou  $m/z_{\text{teoretická}}$  hodnotou

### 2.8.3 Definice jednotky Dalton

Dalton je jednotka molekulové hmotnosti, jedna dvanáctina atomové hmotnosti uhlíku  $^{12}\text{C}$ ,  $1 \text{ Da} = 1,66 \cdot 10^{-27} \text{ kg}$ . Voda tedy má molekulovou hmotnost 18 Da (molární hmotnost 18 g/mol a relativní molekulovou hmotnost 18). Jednotka Da (často se užívají násobky kDa, kilodalton). Jednotka, která nezapadá do systému soustavy jednotek SI, je běžně používána pro vyjádření molekulové hmotnosti biomakromolekul [15].

## 2.9 Možnosti identifikace hmotnostních spekter

Důležitou součástí pro analýzu látek hmotnostní spektrometrií je konečné zpracování naměřených dat. Data, která získáme z MS, jsou reprezentována histogramy rozložení hmotnosti a náboje dané analyzované látky. Při analýzách bohatých směsí látek (např. proteiny a peptidy, nukleové kyseliny) je nutné vzhledem k obsáhlosti získaných dat použít k vyhodnocování počítač. Ten je schopný informace hromadit, uchovávat, analyzovat a hledat souvislosti mezi biologickými daty rychle a efektivně pomocí specializovaného softwaru.

### 2.9.1 Databázové vyhledávací programy

Databázové vyhledávací programy pro identifikaci proteinů z dat získaných MS využívají různé druhy biologických databází. Pracují na principu porovnávání vložených experimentálních dat s teoretickými daty uloženými v těchto databázích. Pokud dojde ke shodě porovnávaných dat na základě určitých kritérií, je protein nebo peptid identifikován s určitým skóre a pravděpodobností a popsán na výstupu z programu [16].

Na internetové adrese <http://www.ionsource.com/links/programs.htm> je uveřejněný přehled volně dostupných programů, které umožňují porovnávání a zpracování naměřených hmotnostních spekter.

### 3 Tandemová hmotnostní spektrometrie

Tandemová hmotnostní spektrometrie je metoda, která podstatným a důležitým způsobem rozšiřuje možnosti klasické hmotnostní spektrometrie. Zahrnuje v sobě vícenásobné kroky hmotnostní analýzy, většinou formou fragmentace.

Tandemová hmotnostní spektrometrie získává v současné době díky bouřlivému rozvoji technologií převahu nad „jednoduchou“ klasickou analýzou. Pro výzkum sekvencí proteinů má pak obzvláště velký význam, protože místo jednoho spektra peptidových iontů, nám umožní získat pro každý peptid hmotnostní spektrum jeho fragmentů.

#### 3.1 Princip metody TMS

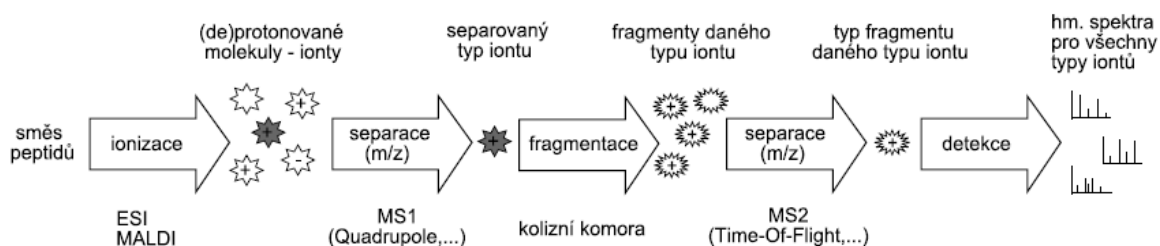
Na rozdíl od jednoduché MS, umožňuje zvolit peptid, který je následně fragmentován kolizí s inertním (netečným) plynem. Profil výsledků fragmentace (*fragmentation pattern*) poskytuje částečnou či úplnou informaci o sekvenci proteinu, která je vodítkem pro hledání shody s daty uloženými v databázích.

#### 3.2 Tandemový hmotnostní spektrometr

Tandemový hmotnostní spektrometr má obvykle dva analyzátory oddělené kolizní komorou (*collision cell*), která je vyplněna inertním plynem (např. argon, xenon). Tandemová hmotnostní analýza pak probíhá ve 2 fázích.

V první fázi jsou všechny ionty podle poměru  $m/z$  postupně vpouštěny do kolizní komory, kde dochází ke kolizně indukované disociaci (*Collision Induced Dissociation, CID*), tj. srážkám s molekulami netečného plynu a rozpadu iontů na fragmenty.

Ve druhé fázi jsou všechny fragmenty daného typu iontu (určeného poměrem  $m/z$ ) opět podle poměru  $m/z$  použity k detektoru. Tímto postupem získáme pro každý typ iontu hmotnostní spektrum jeho fragmentů. (Obrázek 6).



**Obrázek 6:** Tandemová hmotnostní spektrometrie [10]

Získaná MS/MS spektra obsahují pouze fragmentové ionty vzniklé rozpadem daného typu iontu a neobsahují žádné nečistoty. Oba použité analyzátory mohou být stejného typu (*Quadrupole-Quadrupole, TOF-TOF*) nebo různých typů (*Quadrupole-TOF* tzv. *QTOF*), apod.

Moderní spektrometrické metody umožňují nejen  $MS^2$ , ale dokonce i  $MS^n$  hmotnostní analýzu, kde  $n \leq 10$ .

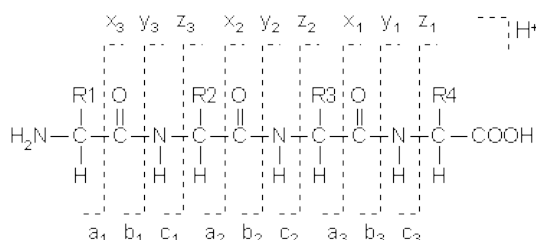
Hmotnostní spektrometrie může být tandemová v čase nebo prostoru. Např. uvnitř iontové pasti probíhá izolace a následná aktivace a fragmentace v jednom místě. Je to tedy tandem v čase, kdy jsou



napřed nežádoucí ionty vypuzeny, a potom jsou ionty s vhodným  $m/z$  fragmentovány srážkami nejen mezi sebou, ale i s atomy helia. Naproti tomu při tandemu v prostoru dochází k selekci, fragmentaci a separaci v různých částech hmotnostního spektrometru. Příkladem takových spektrometrů je *TOF-TOF*. Ten je složen ze dvou *TOF* analyzátorů, mezi nimiž je kolizní cela. Zde však mají ionty při fragmentačních procesech vyšší energii, a proto jsou tyto procesy považovány za vysoko energetickou *CID* (ke fragmentaci často stačí jen jediná srážka)[10].

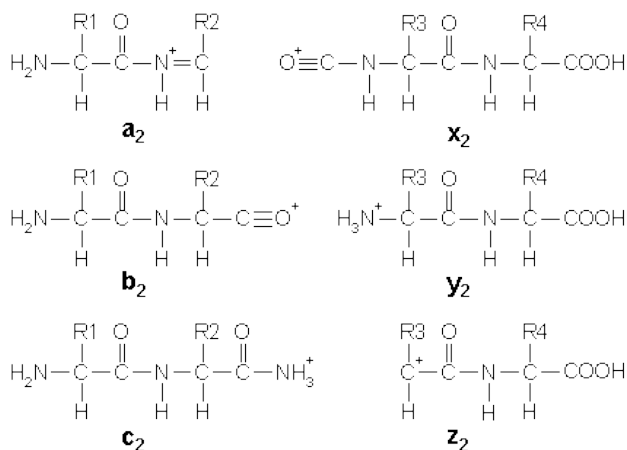
### 3.3 Fragmenty spekter

V dnešní době používané hmotnostní spektrometry umožňují určení hmotnosti peptidu či bílkoviny s velmi velkou přesností. Avšak ani ta obvykle nestačí k jejich jednoznačné identifikaci; i když přesně změřená hmotnost může u menších peptidů odhalit celkové složení, nevypovídá nic o kovalentní struktuře peptidu. Tu lze zjistit sekvenováním polypeptidu (<25 aminokyselin) pomocí tandemové hmotnostní spektrometrie (*MS/MS*) tak, že peptidu, jehož sekvenci chceme určit, je dodána energie, která vyvolá fragmentaci, a následně je změřeno hmotnostní spektrum fragmentů.



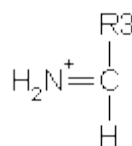
**Obrázek 7:** Určení C- a N- konců peptidu [17]

Protože většina peptidů a proteinů jsou lineární polymery, přerušením jediné kovalentní vazby v řetězci vznikají různé typy iontů v závislosti na místě přerušení peptidového řetězce. Tak vznikají dvě částice, obsahující N- a C-koncovou část peptidu (Obrázek 7). K tomu, aby byla vzniklá částice detekována, musí nést nejméně jeden náboj. Jestliže je náboj zadržen N-koncovou částí peptidu, ion je klasifikován jako **a**, **b** nebo **c**; je-li zadržen C-koncovou částí, je ion klasifikován jako **x**, **y** nebo **z**, a to podle vazby, v níž došlo k fragmentaci. Číslo v dolním indexu udává počet aminokyselin ve fragmentu (Obrázek 8).



**Obrázek 8:** Klasifikace iontů [17]

Při fragmentaci může dojít k přerušení více než jedné peptidové vazby. Tím vznikají tzv. interní fragmenty. Zvláštním typem interních fragmentů jsou immoniové ionty, které obsahují pouze jedinou aminokyselinu (Obrázek 9). Přítomnost těchto iontů ve spektru jednoznačně potvrzuje výskyt dané aminokyseliny v sekvenovaném peptidu.



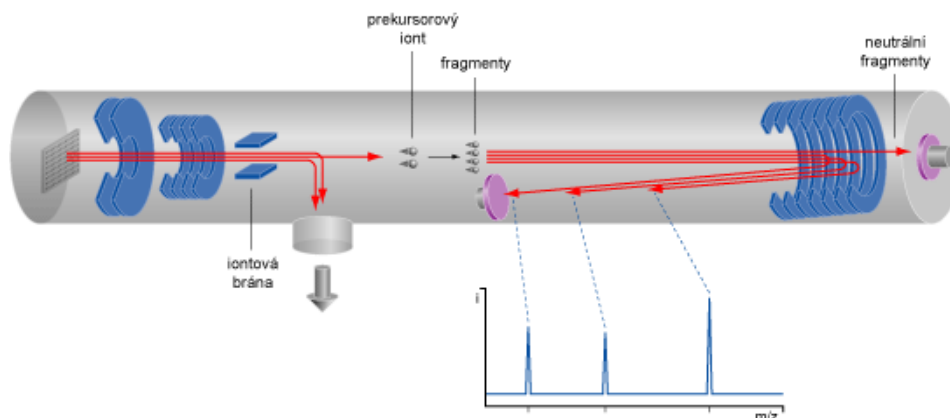
**Obrázek 9:** Immoniový iont [17]

V současné době je nejčastěji používaná fragmentace peptidu metodou tzv. kolizí vyvolané disociace (*collision-induced dissociation*, CID), a to zejména ve spojení s trojitým kvadrupólem nebo TOF-TOF jako hmotnostními analyzátory. V případě trojitého kvadrupólu slouží první kvadrupól k selekci peptidu, který chceme sekvenovat (prekursor, mateřský ion). Druhý kvadrupól je naplněn inertním plynem, nejčastěji argonem. Při srážkách iontů peptidu s molekulami kolizního plynu dochází k přeměně kinetické energie na energii vnitřní a tím k prasknutí některé vazby v peptidu. Hmotnosti nabitých fragmentů jsou pak změřeny třetím kvadrupólem. Podobně funguje TOF-TOF tandemový hmotnostní spektrometr, kde první TOF analyzátor vybírá prekursorové ionty, které vstupují do kolizní cely umístěné mezi TOF analyzátory, a spektrum fragmentů je měřeno druhým TOF analyzátozem.

CID se dá použít i ve spojení s kvadrupólovou iontovou pastí. Nejprve dochází k akumulaci všech iontů přítomných v měřeném vzorku v iontové pasti, poté dojde k vypuzení iontů, jež nebudou fragmentovány. Vnitřní energii nutnou pro fragmentaci získá iont při mnoha srážkách s plynem přítomným v iontové pasti po přivedení rezonančního potenciálu na vstupní a výstupní elektrodu. Na závěr jsou z iontové pasti vypuzeny iontové fragmenty a zaznamenáno jejich hmotnostní spektrum. Kterýkoliv z fragmentů vzniklý během MS/MS experimentu v iontové pasti může být znovu zachycen v pasti, aktivován, fragmentován a může být změřeno hmotnostní spektrum fragmentů tohoto fragmentu; Tento proces může být opakován, dokud se nedosáhne dobrého rozlišení a tím i určení sekvence.

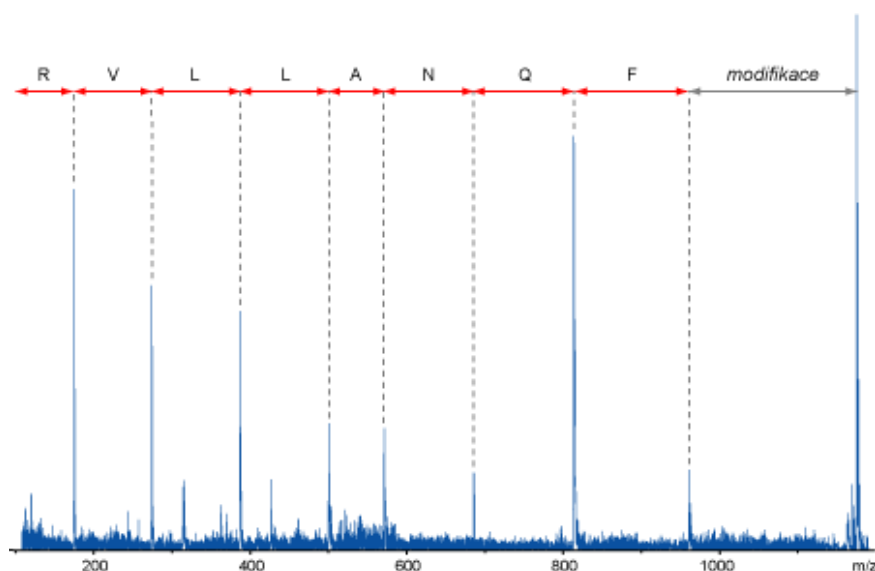
Fragmentaci molekul je vyvolána nejen kolizí již ionizovaných molekul s neutrálními molekulami, ale také vysokou excitací analyzovaných molekul při samotné ionizaci; k rozpadu molekuly na fragmenty tak dochází již v iontovém zdroji (*In Source Decay*, ISD). Tímto způsobem lze analyzovat pouze čistou látku (interpretace spekter, ve kterých by se vyskytovaly fragmenty několika prekursorů, by byla velice obtížná). Při použití ESI lze ISD dosáhnout zvýšeným napětím vloženým na kapiláru, u MALDI zvýšeným výkonem laseru.

Při použití spektrometru na principu MALDI-TOF (Obrázek 10) se pro získání sekvence peptidu využívá rozpadu ionizovaných molekul prekursoru v letové trubici již bez přítomnosti elektrického pole; k rozpadu iontů tedy dochází až za zdrojem (*post-source decay*, PSD). Nutná je však přítomnost iontového selektoru (iontová brána, *ion gate*, deflektor), který ze směsi peptidů (prekursorů) vybere ten, jehož sekvenci chceme určit. K hmotnostní analýze fragmentů je dále třeba, aby MALDI-TOF spektrometr byl vybaven reflektorem, neboť fragmentové ionty mají stejnou rychlost, ale rozdílnou hmotnost. Čím těžší je fragmentový iont, tím má vyšší kinetickou energii, tím hlouběji pronikne do reflektoru a tím delší bude jeho doba letu k reflektorovému detektoru.



**Obrázek 10:** Schéma sekvenování peptidů spektrem MALDI-TOF [18]

Spektra, získaná metodou PSD obsahují převážně a, b a y fragmenty a immoniové ionty jednotlivých aminokyselin. Fragmentace však nebývá úplná, jednotlivé série iontů jsou nekompletní a tím je interpretace spekter poměrně obtížná. (Obrázek 11) [18].



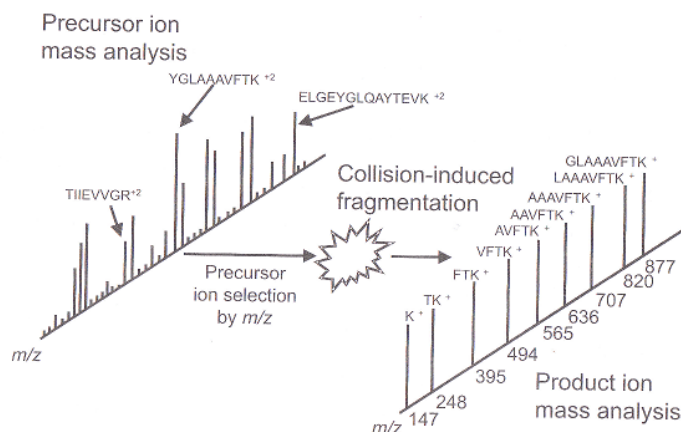
**Obrázek 11:** PSD MALDI-TOF spektrum peptidu [18]

### 3.4 Výsledky z MS/MS

#### 3.4.1 Rozdíl mezi výsledky z MS a MS/MS

Hmotnostní analýza (MS) je v podstatě separace iontů podle jejich  $m/z$ . Tandemové hmotnostní spektrometry používají tohoto dělení jako přípravný nástroj k dělení iontů s určitým  $m/z$  pro další analýzu. Tato další analýza se provádí hmotnostním tříděním vybraných iontů a stanovením  $m/z$  fragmentu iontů v druhé fázi celkové analýzy. Termín "tandemová hmotnostní spektrometrie" odráží skutečnost, že dva ionty v složité směsi by měly být selektivně studovány v experimentu. Z výsledků

je patrné to, že konkrétní ionty ve složité směsi mohou být selektivně studovány v experimentu, který poskytuje informace o struktuře iontů. V případě peptidu iontů, strukturální informace je amino kyselina dané sekvence peptidu. Obrázek (Obrázek 12) ilustruje tento postup pro charakterizaci jedné hmotnosti peptidu, vybraných iontů v hmotnostním spektru, které obsahuje několik peptidů [19].



**Obrázek 12:** Strukturní charakterizace hmotnosti vybraného iontu pomocí MS/MS [19]

### 3.5 Přehled možností identifikace MS/MS dat

Nyní se budeme zabývat programovými možnostmi identifikace získaných dat po vlastním fyzickém procesu rozpoznávání analytu. Data získaná z MS/MS jsou reprezentovány histogramy rozložení hmotnosti analyzované látky. Spektrum výsledných relativních molekulových hmotností je porovnáváno se soubory hmotností, které byly vytvořeny počítačem z proteinů v proteinové databázi.

Existuje celá řada komerčních i volně dostupných algoritmů, které byly vyvíjeny od poloviny devadesátých let a jejich vývoj pokračuje i nyní. Tyto algoritmy se podle svých vlastností a podle schopnosti rozpoznání dělí do tří základních skupin.

#### 3.5.1 „De Novo“ peptidové sekvenování

Metoda je založena na přímém interpretování tandemového hmotnostního spektra, pro samotné určení sekvence peptidů využívá grafové algoritmy. Výhodou této metody je to, že dokáže identifikovat i sekvence, pro které se doposud nenašlo místo v žádné databázi. Nevýhodou je naopak to, že počet řešení může být velmi vysoký, protože danému spektru může odpovídat mnoho peptidových sekvencí. Technika „De Novo“ je v praxi proto velmi neefektivní a její běžně dostupné implementace dovedou správně identifikovat méně než 30% sekvencí [21].

Přehled několika známých implementací je uveden v následující tabulce.

**Tabulka 1:** Přehled dostupných implementací pro metodu "De Novo"

Název	Webová adresa
Lufetisk	<a href="http://www.hairyfatguy.com/lutefisk/">http://www.hairyfatguy.com/lutefisk/</a>
Peaks	<a href="http://www.bioinformaticssolutions.com/products/peaks/index.php">http://www.bioinformaticssolutions.com/products/peaks/index.php</a>
Pepnovo	<a href="http://proteomics.bioprojects.org/Software/PepNovo.html">http://proteomics.bioprojects.org/Software/PepNovo.html</a>

Tyto implementace obvykle nevyužívají přesnou hmotnost aminokyselin a iontů, ale berou ohled na chybovost a pracují i s určitými odchylkami hodnot.

### 3.5.2 Identifikace s využitím tagů

Metoda Identifikace s využitím tagů (*Sequence Tag*) se snaží kombinovat výhody přímé interpretace spektra a vyhledávání v databázích. Z daného spektra peptidu se nejprve podle vzdáleností píků snažíme vytipovat část sekvence (tag) a následně provádíme dohledání celých proteinových sekvencí v databázi. Pro definování tagu můžeme využít postupy založené na metodě „De Novo“ nebo jej můžeme určit i „ručně“ na základě vizuální interpretace části sousedících píků. Právě kvůli snadným manuálním úpravám, které mohou výrazně zlepšit kvalitu výstupu, je tato metoda poměrně populární [22].

Přehled několika veřejně dostupných webových aplikací podporujících metodu Sequence Tag:

**Tabulka 2:** Přehled dostupných implementací pro metodu "Sequence Tag"

Název	Webová adresa
Mascot	<a href="http://www.matrixscience.com/search_form_select.html">http://www.matrixscience.com/search_form_select.html</a>
PeptideSearch	<a href="http://www.narrador.embl-heidelberg.de/GroupPages/PageLink/peptidesearchpage.html">http://www.narrador.embl-heidelberg.de/GroupPages/PageLink/peptidesearchpage.html</a>
ProteinProspector MS-Seq	<a href="http://prospector.ucsf.edu/">http://prospector.ucsf.edu/</a>

### 3.5.3 Metoda fragmentového mapování

Při použití metody fragmentového mapování (*Peptide Fragment Fingerprinting*, PFF) se porovnává experimentální spektrum peptidů s teoretickými spektry generovanými z databáze proteinových sekvencí. Protože místo jednoho spektra je k dispozici celá sada fragmentových spekter, je identifikace tímto způsobem mnohem přesnější. Čím více spekter se podaří spárovat, tím je pravděpodobnější, že analyzovaný protein odpovídá dané sekvenci [20].

V současné době existuje řada aplikací využívajících pro identifikaci peptidů neinterpretovaná tandemová spektra. Základní vlastnosti jsou uvedeny v příloze (Příloha 1).

## 3.6 X!Tandem

Projekt organizace *The Global Proteome Machine Organization* X!Tandem je volně šiřitelný „open source“ program, který byl optimalizován k použití na méně výkonných výpočetních zdrojích. Kód byl vytvořen v programovacím jazyce C++ pomocí „Standard Template Library“.

Pro X!Tandem je velmi důležitý předpoklad:

*„Pro každý identifikovatelný protein v původní směsi proteinů, bude existovat alespoň jeden zjištěný peptidový list (tryptic peptide) s nulovým nebo jedním vynechaným místem štěpení.“*

Rafinované nebo sekundární analýzy vyvozující bílkovinné položky jsou pak podrobněji analyzovány (podobně jako chyba, která je tolerantní k vyhledávání) s přihlédnutím k nespecifické hydrolýze a / nebo k PTMs.

Sadu tříd, které plní tyto úkoly, tvoří:

1. čtení XML vstupní parametrů souboru,
2. čtení bílkovin z FASTA souborů
3. čtení MS/MS spekter v běžných formátech ASCII (DTA, PKL, MGF),
4. stav MS/MS spektra k odstranění šumu a společných artefaktů,
5. peptid se sekvencí je podroben procesu štěpení činidla, post- translačních a chemických změn,
6. sekvence peptidu skóre,
7. vytvořit výstupní XML soubor, který zachytí nejlepší bodování sekvence a některé statistické rozdělení důležité pro bodování procesu.

X! Tandem je více-vláknový program, který využívá možnosti násobení procesoru do jediného počítače. Počet upravených vláken lze jednoduše změnit v parametru vstupního XML souboru v závislosti na hardwarových prostředcích, které jsou k dispozici. Větší hmotnostní spektra jsou rozdělena na clusteru tak, aby rychlost zpracování velkého MS/MS datového souboru nebyla zmenšena. V poslední době byl upraven X!Tandem kód tak, aby byly skórovány nové funkce, které mají být zapojeny s cílem usnadnit bodování přístupy.

Aktuální verze X!Tandem bodování funkce počítá počet bodů na základě dot-produkt mezi teoretickým (pouze b a y ionty) a experimentálním tandemovým hmotnostním spektry. Skóre je následně převeden do očekávané hodnoty (*E-value*, E-hodnota). E-hodnota představuje počet peptidů v databázi, u kterých se očekává, že dosáhne tohoto skóre sám náhodně (náhodné shody). Poté platí následující pravidlo:

*„Čím nižší hodnota E-value, tím výraznější skóre.“*

E-hodnota se získá při sběru statistických údajů při vyhledávání pro odhad distribuce. U této distribuce se předpokládá, že je hypergeometrická, což je diskrétní rozdělení pravděpodobnosti a proto může být vysoká E-hodnota bodování peptidů získána extrapolací.

Výhodou X!Tandemu je jeho rychlost, v násobcích je přibližně 200x rychlejší pro blíže nespecifikované vyhledávání a cca 1000x rychlejší pro specifické (například pro oxidaci). Další výhody jsou schopnosti pracovat i se semi-tryptidovými peptidy a polymorfními sekvencemi. X!Tandem také využívá ohodnocení (skóre) založené na pravděpodobnosti odhadu správného výsledku.

Současná omezení algoritmu X!Tandem je neschopnost vyhledávat více změn na stejném zbytku kyseliny (např. methyl-a dimethyl arginin) [20].

### 3.6.1 Pracovní postup

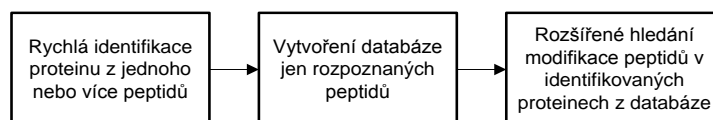
X!Tandem Workflow. X!Tandem porovnává každé spektrum se všemi možnými kandidáty peptidů, které jsou uloženy v databázi proteinů.

Před samotným počátkem vyhledávání jsou sekvence proteinů rozbaleny do peptidových listů (*peptide tryptic*) a pro každé ionty tohoto peptidu je vypočtena hmotnost jejich fragmentů. Tyto potom vytvářejí sestrojené MS charakteristické spektrum hmotností, nazývajícím se *mass fingerprint*, nebo peptidová mapa konkrétního peptidu, který je měřen [20].

Srovnávací proces je pomocí programu X!Tandem rozdělený do dvou kroků sekvence:

1. Zpřístupnění kompletní databáze spekter proteinů, která umožňuje téměř okamžité vyloučení neshodných sekvencí a určení sady proteinů jako možných kandidátů.
2. Na kandidátech se provádí tzv. čisté hledání a identifikace.

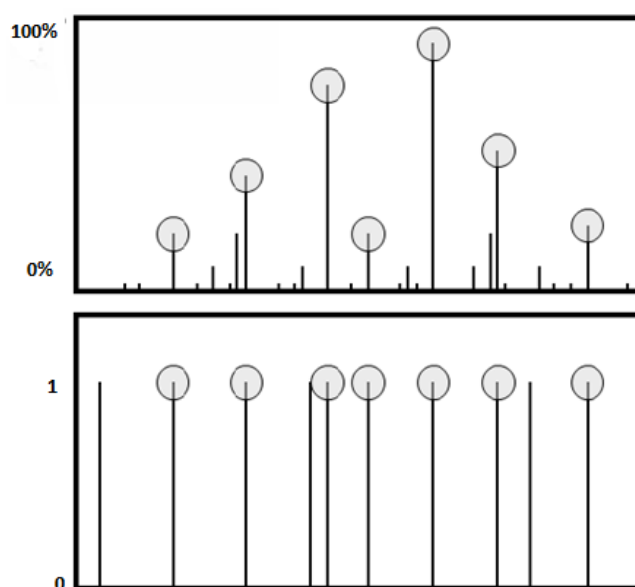
Pracovní tok (*workflow*) algoritmu můžeme shrnout do následující posloupnosti (Obrázek 13):



**Obrázek 13:** Pracovní tok algoritmu

### 3.6.2 Schéma skórování algoritmu X!Tandem

X! Tandem porovnává odpovídající získaná MS / MS spektra s modelovým (vzorovým) MS/MS spektrem na základě peptidů v databázi bílkovin. Modelová (vzorová) spektra jsou velmi jednoduchá, tzn., jsou založena na přítomnosti či nepřítomnosti píku iontů (konkrétně pro typ iontů „y“ a „b“).



**Obrázek 14:** Srovnání hmotnostních spekter [26]

V úvahu jsou brány pouze odpovídající spektrální špičky (označené na obrázku). Jakékoli vrcholy, které se neshodují, a to buď modelově, nebo získaným spektrem, nejsou používány.

Získané spektrum je zjednodušeno pouze na ty vrcholy, které jsou podobné vrcholům v modelovém (vzorovém) spektru.

**Předběžné hodnocení  $S_{y/b}$**  (skóre) je výsledek skalárního součinu změřeného a modelového (vzorového) spektra. Po provedené filtraci všech srovnaných spekter se intenzity překrývajících se piků iontů sumují.

$$S_{y/b} = \sum_{i=0}^n I_i \cdot P_i \quad (1)$$

kde „ $S_{y/b}$ “ je předběžné skóre, „ $I_i$ “ je intenzita naměřeného iontu reprezentovaným píkem v grafu a „ $P_i$ “ je absence (0), nebo presence (1) vzorového píku.

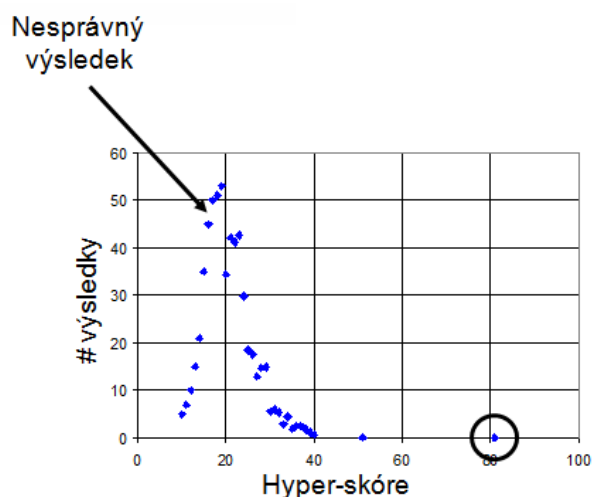
**Hyper-skóre** označuje upravené hodnocení shody v X!Tandemu. Předběžné skóre se upravuje vynásobením  $N!$ , faktoriál pro počet „b“ a „y“ měřených iontů. Použití faktoriálu je založen na hypergeometrickém rozdělení. Hyper-skóre se vyjadřuje pomocí písmene  $H$ .

$$H = (\sum_{i=0}^n I_i * P_i) * N_b! * N_y! \quad (2)$$

kde „ $N_b$ “ je počet „b“ iontů, „ $N_y$ “ je počet „y“ iontů, „ $I_i$ “ je intenzita (velikost) i-tého píku, „ $P_i$ “ je pravděpodobnost výskytu i-tého píku, ta se pohybuje v rozsahu  $<0,1>$ .

### 3.6.3 Histogram hyper-skóre

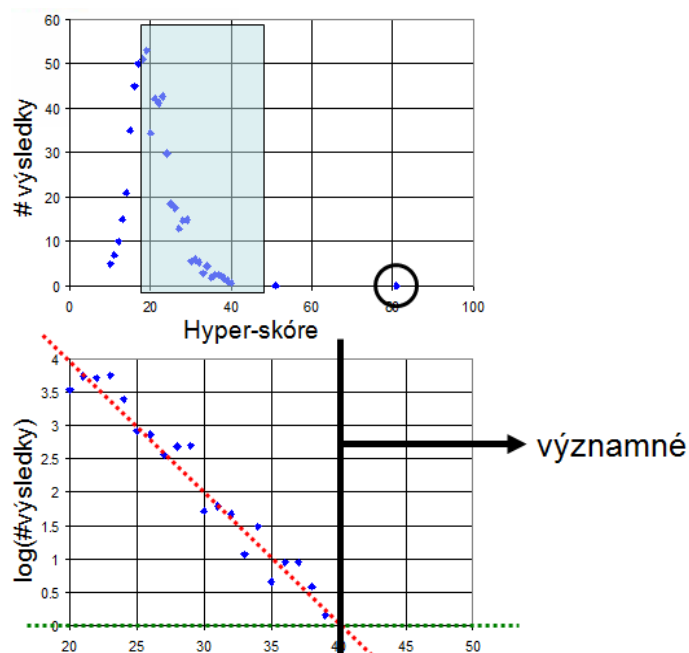
Další krok X!Tandemu spočívá ve vytvoření histogramu hyper-skóre všech spekter peptidů v databázi, která by mohla odpovídat této databázi kandidátů (Obrázek 15). X!Tandem předpokládá, že peptid s nejvyšším hyper-skóre je jediný správný (a s tímto bude dále pracovat) a všechny ostatní s menším skóre jsou nesprávné (nevýznamné).



**Obrázek 15:** Výběr hyper-skóre [26]

Následujícím krokem je transformace údajů v pravé části histogramu (barevná část grafu) aplikací logaritmu na přímku (Obrázek 16). Přímku tvoří očekávané výsledky ze statistického argumentu, které dokazují, že nesprávné výsledky jsou náhodné.



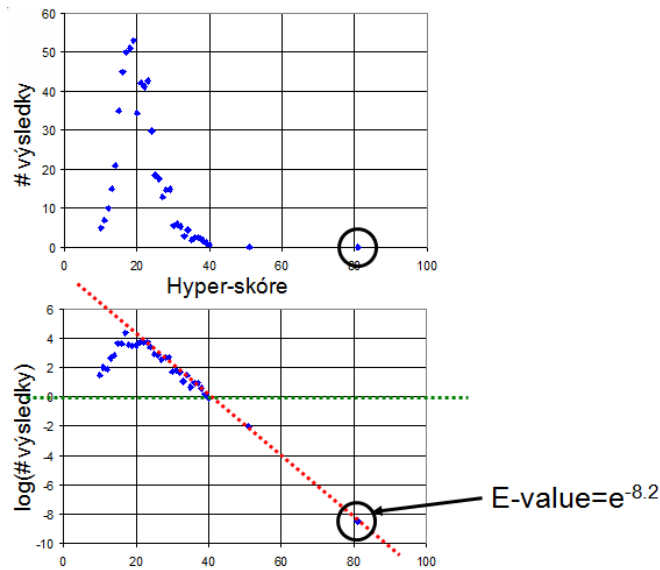


**Obrázek 16:** Určení hranice statistické významnosti H [26]

X!Tandem tedy správně předpokládá, že nejvyšší hyper-skóre je jediný možný správný výsledek nalezení shody (*match*) a tím určení proteinu. Tento výsledek je významný, jestliže je větší než bod, ve kterém se přímka protíná s nulovou osou y a logaritmus výsledků hyper-skóre je roven nule. Jakákoli hyper-skóre větší než toto jsou nepravděpodobné tomu, že vznikla náhodou [26].

### 3.6.4 Výpočet pravděpodobnostního koeficientu shody (E-value)

E-value je parametr, který popisuje, jaký počet přístupů lze „očekávat“, aby byla rozpoznána náhoda při prohledávání databáze o určité velikosti. Klesá exponenciálně se skóre (S), které je přiřazeno ke shodě (*match*) mezi dvěma sekvencemi. E-value v podstatě popisuje náhodný šum pozadí, která existuje pro shodu mezi sekvencemi. Když je k hitům přiřazena hodnota E-value „1“, lze očekávat, že v databázi dané velikosti, se vyskytuje shoda sekvencí s podobným skóre jen náhodou. To znamená, že čím nižší je E-value, nebo je blíže k „0“, tím vyšší je význam této shody. Je však důležité si uvědomit, že vyhledávání v krátkých sekvencích může být prakticky identické a má relativně vysoké E-value. Je to proto, že výpočet E-value také bere v úvahu dotaz na délku sekvence, protože kratší sekvence mají vysokou pravděpodobnost toho, že se vyskytují v databázi čistě náhodou.



**Obrázek 17:** Grafické určení E-value [26]

Výpočet pravděpodobnostního koeficientu shody (*expectation value*) vzorového peptidu (jeho spektra) s peptidem z naměřených hodnot (jeho spektra) je nepoužívanější metodou X!Tandemu.

Ze všech hodnocených spekter je pro každý peptid vypočítán histogram hyperskóre. Nejvíce hodnocená spektra jsou pak shodná a validní. Do náhodných shod jsou pak zahrnuta ostatní spektra. Vyjadřuje pravděpodobnost validního skóre, tzn. pravděpodobnost, že skóre je náhodné a odvozeno z logaritmicko-lineární extrapolace pravé části distribuce hodnot. Násobením této hodnoty počtem hodnocených sekvencí udává hodnotu ohodnocení (*expected number*). V této hodnotě je zahrnut peptidový list a daná sada spekter. Jakmile je určeno peptidové složení, X!Tandem přechází k určení proteinu. To je založeno na počtu správně určených peptidů „*n*“ (*hits*) v proteinu a jejich partikulárních ohodnocení  $e_i$ .

Rovnice je bayesovský model pro proteiny se shodami o různých hodnotách pravděpodobnosti. První dva vztahy v rovnici popisují pravděpodobnost náhodného původu hmotnostních spekter.

Základem je generování hmotnostního spektra „*s*“. Jestli je proteinová sekvence, která je usuzovaná z „*n*“ peptidových sekvencí (jedinečných), každá obdrží hodnotu pravděpodobnosti „ $e_j$ “. Vypočtená hodnota pravděpodobnosti celkové shody (*expectation*) pro hledaný protein je označena  $E_T$  a je vypočítána jako:

$$E_T = \binom{s}{n} \cdot \left( \frac{\beta^n (1-\beta)^{s-n}}{s N^{n-1}} \right) \cdot \left( \prod_{j=1}^n e_j \right) \quad (3)$$

$$E_T = \left( \prod_{j=0}^{n-1} \frac{\binom{s-i}{n-i}}{\binom{s-i}{n-i}} \right) \cdot \left( \frac{\beta^n (1-\beta)^{s-n}}{s N^{n-1}} \right) \cdot \left( \prod_{j=1}^n e_j \right) \quad (4)$$

Kde význam jednotlivých členů rovnice je následující:

- „ $E_T$ “ je celková pravděpodobnost shody pro analyzovaný protein.
- „ $e_j$ “ je pravděpodobnost shody pro  $j$ -tou sekvenci.
- „ $n$ “ je celkový počet jedinečných sekvencí.
- „ $s$ “ je spektrum molekulových hmotností (pořadí v sadě dat).
- „ $N$ “ je počet peptidových sekvencí nalezených peptidů (jedinečných)
- „ $i$ “ je index pořadí.
- „ $j$ “ je index pořadí.
- „ $\beta$ “ je normalizovaný počet peptidů.

Hodnota  $\beta$  se pohybuje v intervalu  $<0,1>$  a je vypočtena ze vztahu:

$$\beta = \frac{N}{\text{celkový počet peptidů v uvažovaném proteomu}} [-] \quad (5)$$

Na celkový výsledek pravděpodobnostní shody „ $E$ “ pro protein, mají vliv dílčí hodnoty pravděpodobnosti shody pro jednotlivé peptidy „ $e_j$ “, z nichž se protein skládá. Hodnoty peptidů jsou kombinované jednoduchým Bayesovským modelem pro klasickou pravděpodobnost. Ve speciálním případě, kdy je pozorovaný jen jeden peptid, se vzorec zjednoduší na tvar:

$$E_T = e_1 \quad (6)$$

Hodnota expectation ( $E_T$ ) tedy vyjadřuje to, jak nepravděpodobné je výsledné hyperskóre pro celý protein (čím nižší, tím lepší). Čím je hodnota expectation vyšší, tím je větší možnost, že náhodný. X!Tandem spočítá hodnotu  $E_T$  pomocí extrapolace přímky logaritmu histogramu. Čím vyšší hodnota expectation, tím se dá méně věřit vypočtenému hyperskóre. Žádané jsou tedy proteiny s nízkou hodnotou expectation [27].

### 3.6.5 Binomické rozložení

#### Výpočet očekávané hodnoty bílkovin z očekávaných hodnot peptidů v X!Tandem

Na rozdíl od některých předchozích generací vyhledávacích algoritmů, všechny série X! počítají statistické hodnoty pro všechna jednotlivá spektra ze sekvence pomocí seřazení. Dále spojují seřazení všech peptidů do souboru údajů do známých sekvencí bílkovin a přiřazují jim statistickou jistotu, že toto kompletování a vyrovnání je nenáhodné.

Předpokládáme, že experimenty přinesly  $s$  hmotnostních spekter. Jestliže proteinové sekvence jsou odvozeny z pozorování  $n$  unikátních peptidových sekvencí, z nichž každý má přiřazenou očekávanou hodnotu  $e_j$ , pak očekávaná hodnota pro bílkoviny  $E_{pro}$  je dána vzorcem:

$$E_{pro} = \left( \frac{\beta^n (1-\beta)^{s-n}}{sN^{n-1}} \right) \times \left( \prod_{j=1}^n e_j \right) \times \left( \prod_{i=0}^{n-1} \frac{(s-i)}{(n-i)} \right) \quad (7)$$

kde  $N$  je peptidová sekvence skórování v  $n$  nalezených unikátních peptidech,  $\beta = N/(\text{celkový počet peptidů v posuzovaném proteomu})$ .

Ve výjimečném případě, kdy byl pozorován pouze jeden peptid  $E_{pro} = e_1$  [30]

Hodnota expectation ( $E_{pro}$ ) vyjadřuje nepravděpodobnost výsledného hyperskóre pro celý protein. Čím je tato hodnota nižší (menší), tím je zmenšena možnost, že je náhodný. Hodnota  $E_{pro}$  je v programu

X!Tandem vypočítána pomocí extrapolace přímky logaritmu histogramu. Žádané jsou proteiny s nízkou hodnotou expectation [31].

### 3.7 OMSSA

*Open Mass Spectrometry Search Algorithm* je stejně jako X!Tandem volně šiřitelný „open-source“ vyhledávací algoritmus, který byl vyvinut v *National Center for Biotechnology Information* (NCBI).

OMSSA může být použita buď online, s omezením velikosti předložených vstupních dat, nebo ji lze stáhnout a nainstalovat lokálně na Windows, Mac OS X nebo Linuxu. Vyhledávací algoritmus byl nedávno použit jako primární nástroj pro identifikaci charakterizující lidského slinného proteomu [20].

Velké množství MS/MS spekter peptidů získaných pomocí proteomických experimentů vyžaduje efektivní, citlivé a specifické algoritmy pro identifikaci peptidů. OMSSA vypočítává specifičnost klasickým pravděpodobnostním skóre pomocí explicitního modelu pro odpovídající experimentální spektra se sekvencí.

OMSSA je unikátní svým použitím klasického testování hypotéz na základě výslovného modelu odpovídající statistiky, typu statistických modelů používaných v BLAST. Vzhledem k velkému počtu získaných spekter ve vysoce propustné (výkonné) proteomice, je manuální interpretace nepraktická, a proto je důležité, aby shody (match) těchto algoritmů skórovaly s prahem nebo prahy, které umožňují několik falešných výsledků. Pravděpodobnost vyrovnání umožňuje, aby takové limity, které mají být stanoveny z hlediska počtu falešných výsledků, měly povoleno vést k vývoji několika statisticky vyhodnocených algoritmů vedle stávajících na základě pravděpodobnosti vyhledávacích algoritmů.

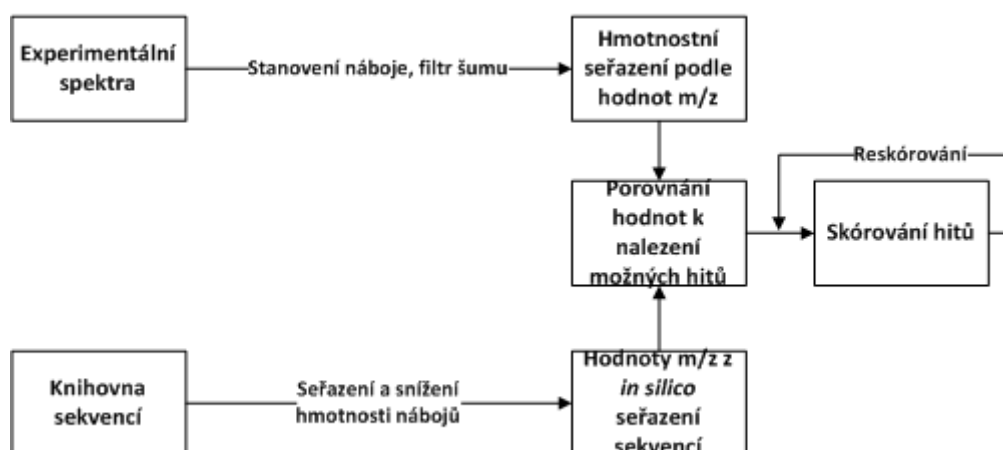
OMSSA se pokouší vytvořit rychlý vyhledávací algoritmus, jehož výsledky jsou skórovány za použití klasického statistického modelu, pomocí předpokladů převzatých přímo od experimentálního nastavení a umožňuje experimentální šum.

#### 3.7.1 Pracovní tok algoritmu OMSSA

Algoritmus OMSSA pracuje podle následujícího vývojového diagramu (Obrázek 18). Jednotlivé bloky v diagramu jsou popsány dále. Pro snadnější popis algoritmu použijeme tři teoretické produkty iontů.

##### Určení náboje

MS/MS spektra často neobsahují experimentálně určená spektra z citlivostních důvodů. Nicméně zjištění možného prekurzoru náboje spektra je nezbytné, protože OMSSA vybírá teoretické peptidy z knihovny sekvencí tím, že je porovná s neutrální hmotností prekurzoru. K tomu také OMSSA určuje, zda je nebo není daný prekurzor prvního náboje, tím že spočítá počet píků nad prekurzorem z hodnoty  $m/z$ .



**Obrázek 18:** Vývojový diagram algoritmu

### Filtrace šumu

Experimentální spektra mohou obsahovat značné množství šumu, které může mít za následek náhodné shody (*matches*), které vyžadují algoritmus pro odstranění píků šumu bez odstranění píků signálu.

#### Kroky filtrace šumu:

- Algoritmus OMSSA před vstupními procesy předpokládá odstranění vrcholů spektra, které představují pozadí. Vrcholy by měly být odstraněny, pokud jejich míra je nižší než 2,5% maximální intenzity. (Tato hodnota je uživatelsky nastavitelná a dynamicky se mění v poslední části algoritmu.)
- Následné kroky zahrnují odstranění veškerých píků, které nejsou považovány za monoizotopické, nebo píky, které jsou příliš blízko nejčastějším píkům [20]. Toho je dosaženo tím, že píky jsou zkoumány v pořadí podle intenzity a jsou odstraněny vrcholy, které jsou v rozmezí 0 - 2 Da z hodnoty m/z píku zkoumaného.
- Posledním krokem v hluku filtru je odfiltrovat vrcholy, které jsou příliš blízko u sebe.

### Výpočet a porovnání s prekurzorem hmotnosti

První srovnání je provedeno mezi experimentálním spektrem a knihovnou hledání pořadí, tím že jsou porovnány naměřené prekurzory a hmotnosti jsou stanoveny výpočtem *in silico*. Pokud je shoda v toleranci, kterou stanovil uživatel, algoritmus postupuje do dalšího kroku, jinak jsou vybrány další peptidové sekvence z databáze a srovnány s prekurzorem hmotnosti. Výpočet teoretické hmotnosti peptidu umožňuje minout výstřih, změny hmotnosti aminokyselin a variabilní úpravy hmotnosti aminokyselin, kde "proměnná" znamená, že hmotnosti jsou vypočteny s modifikací a bez modifikace. Tento krok je výpočetně nejnáročnější [29].

V zájmu urychlení algoritmu, se pro výpočet používá několik strategií:

1. Výpočet je proveden pomocí celých čísel. Navýšení hodnoty může být upraveno uživatelem ve zdrojovém kódu.
2. Databáze sekvencí je ukládána do paměti, aby mohlo dojít k rychlému načítání sekvence do procesoru.
3. Spektra vyhledávání jsou tříděny a indexovány podle prekursoru hmotnost, aby se zabránilo zbytečnému srovnání.

### **Výpočet hmotnostního žebříčku**

Jestliže hmotnost prekursoru odpovídá hmotnosti vypočtené, teoretické hodnoty  $m/z$  iontu jsou vypočteny z databáze peptidů, aby mohlo dojít k srovnání s hodnotami  $m/z$  odvozenými z experimentálního spektra. B a y iontové řady prvního vzorku jsou vypočtené z náboje 1. a 2. prekursoru iontů. B a y iontové řady druhého vzorku jsou vypočítány z náboje 3. prekursoru iontů.

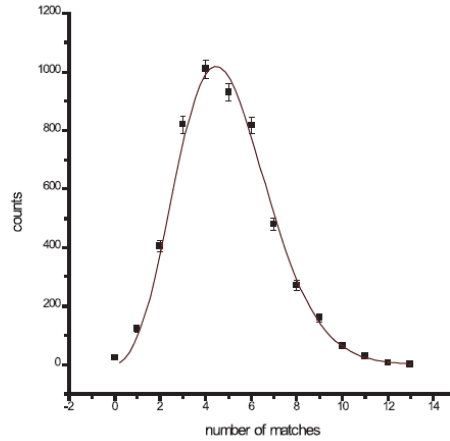
### **Srovnání hmotnostních žebříčků**

K nalezení hitů srovnává algoritmus OMSSA vypočtené hmotnostní žebříčky s hmotnostními žebříčky z experimentálních spekter pomocí hmotnostní tolerance, kterou nastavuje uživatel. Oba hmotnostní žebříčky jsou řazeny podle hodnoty  $m/z$ . Je-li experimentální hodnota poměru  $m/z$  použita při shodě mezi danou experimentální hmotností žebříčku a vypočtenou hmotností žebříčku, již se nebere v potaz v dalších iontových řadách, např. odpovídá-li experimentální hodnota  $m/z$  konkrétním b iontům, není dovoleno, aby odpovídala i y iontům. Důvodem tohoto omezení je, že pokud hodnotě  $m/z$  odpovídá více než jeden druh iontů, pak konzervativní přístup předpokládá, že se bude shodovat jen jeden. Uzavřené ionty jsou vloženy do seznamu a jsou seřazeny podle počtu shod. K šetření paměti systému je délka tohoto seznamu omezena, většinou na nastavitelnou hodnotu (100) ve výchozím nastavení [29].

### **Skórování**

Pro zjištění bodování funkcí je důležité pochopit charakteristiky náhodných shod hodnot  $m/z$  odvozených ze spektra. Umožnění výpočtu rozdělení náhodných shod významného nálezu musí být vyjádřeno jako náhodná pravděpodobnost nálezu, kde nízká pravděpodobnost znamená významný zásah.

Obrázek 19 zobrazuje histogram pro jedno ze spekter, které počítá počet produktů hodnot  $m/z$  v každém teoretickém peptidu, který má shodné produkty hodnoty  $m/z$  experimentálního spektra, kdy předchůdce hmotností experimentálního spektra odpovídá vypočtenému prekursoru hmotnosti. Poissonova distribuce je přizpůsobena histogramu, který se nachází v distribuci náhodných procesů, kde je průměrný počet úspěchů mnohem nižší, než možný počet úspěchů.



**Obrázek 19:** Histogram spekter [29]

Nechť  $o$  je spodní hranice měřeného produktu iontů hodnoty  $m/z$  a  $r$  je horní hranice. Pokud je míra tolerance produkovaných iontů  $t$ , pak míra množství možných shod je  $(r-o)/2t$ . Je-li  $m$  neutrální hmotnost předchůdce, pak musí platit, že  $h \cdot (r-o)/m$  se počítá z hodnoty  $m/z$  v experimentální produkci iontů, kde  $h$  je celkový počet vypočítaný z  $m/z$  hodnot produktů ionty [29].

Za předpokladu Poissonova procesu to znamená

$$\mu_1 = \left( \frac{2t}{(r-o)} \right) \cdot \left( \frac{h \cdot (r-o)}{m} \right) \cdot v = \frac{2thv}{m} \quad (8)$$

Pro Poissonovu distribuci

$$P(x, \mu) = \frac{\mu^x}{x!} e^{-\mu} \quad (9)$$

Kde  $x$  je počet měřených shod.

Dále jsou vzaty v úvahu spektra, které obsahují 1. a 2. produkty iontů. Rozdělení pravděpodobnosti je rovněž Poissonova distribuce, jejíž průměr je dán

$$\mu_2 = \mu_1 \cdot \frac{r+m-3o}{r-o} \quad (10)$$

### Výběr $m/z$ hodnot odpovídajících nejintenzivnějším vrcholům

Ke zvýšení účinnosti a citlivosti algoritmus platí pro výběr z teoretických spekter:

- Alespoň jedna z hodnot  $m/z$  teoretických spekter musí odpovídat hodnotám  $m/z$   $n$  nejvyšších vrcholů ve spektru ( $n = 3$  ve výchozím nastavení). Tato volba změní rozdělení pravděpodobnosti.
- Je-li pravděpodobnost, že  $q$  je hodnota  $m/z$  shod vypočtená z  $m/z$  hodnot vztahu  $n/v$ , pak je rozdělení pravděpodobnosti

$$P'(x, \mu) = \frac{1}{q} (1 - (1 - q)^x) P(x, \mu) \quad (11)$$

kde normalizační faktor  $Q$  je

$$Q = \sum_x (1 - (1 - q)^x) P(x, \mu) \quad (12)$$

### 3.7.2 Výpočet E-value

E-value je parametr, který je očekávaným počtem náhodných hitů z hledané knihovny daného spektra tak, že náhodné hity mají stejné nebo lepší skóre než daný hit. Například, hodnota E-value hitu „1,0“ znamená, že jeden jiný hit ze skóre je stejný nebo lepší než hit, který je náhodně vyhledaný z knihovny sekvenční. Je-li pravděpodobnost, že jediné srovnání spektra, aby se vypočítala MS/MS spektra, není náhodné, je

$$\sum_{x=0}^{y-1} P'(x, \mu_z), \quad (13)$$

kde  $y$  je počet úspěšných shod (match) iontů a  $z$  je rovno „1“ nebo „2“ v závislosti na hledaných iontových řadách, pak pravděpodobnost, že  $z$  jednoho spektra proti  $N$  teoretických spektrech je náhodné, je

$$1 - (\sum_{x=0}^{y-1} P'(x, \mu_z))^N \quad (14)$$

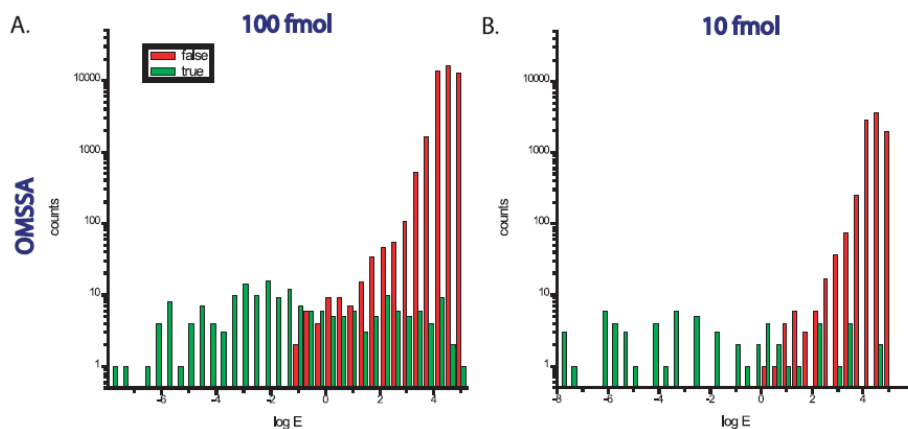
Hodnota E-value je potom

$$E(y, \mu) = N - (1 - (\sum_{x=0}^{y-1} P'(x, \mu_z))^N) \quad (15)$$

Tato hodnota E-value platí pro vyhledávání, které jsou variabilní post-translačními modifikacím. Obecně platí, že proměnná post-translačních úprav navýší hodnoty  $N$ , protože modifikované peptidy mohou vytvářet dvě nebo více teoretických spekter, v závislosti na počtu míst, které jsou modifikovány. Teoretičtější spektra generovaná z jediné peptidové sekvenční mohou být považována za nadbytečné, neboť nemají stejné předchůdce pro výpočet  $m/z$  a sdílí pouze podmnožinu produktu iontů, takže je zbytečné, aby byl prozkoumán dopad nadbytečných hodnot na E-value.

Na obrázku (Obrázek 20) jsou znázorněny histogramy falešně pozitivních a pravdivě pozitivních výsledků. Grafy A a B jsou vytvořeny pro bílkovinné standardy o velikosti 100 fmol a 10 fmol, které byly analyzovány algoritmem OMSSA. Červená barva označuje falešně pozitivní výsledek a zelená pravdivě pozitivní výsledek. Na vodorovné ose je logaritmus E-value a na svislé ose počet přístupů s daným skóre. Hodnoty E-value pod 0,1 jsou považovány za významné [29].





**Obrázek 20:** Histogramy falešně pozitivních a pravdivě pozitivních výsledků [29]

### 3.7.3 Reskórování ke zlepšení citlivosti

Citlivost OMSSA může být zlepšena díky změně hranice používané v počáteční fázi odfiltrování šumu od pozadí. Toho je dosaženo tím, že se změní práh pozadí o 0-20% maximální intenzity vrcholu a E-value prozkoumá nejlepší hit. Zvolený práh je jeden výsledek v nejnižší E-value z nejlepšího hitu. Je možné, že toto reskórování může ovlivnit nepříznivě statistiku převzorkováním sekvence knihovny, ale v praxi to výrazně nezmění seznam nejlepších hitů [29].

## 4 Datové struktury

Na internetu je uživatelům zpřístupněno velké množství zdrojů dat. Databáze obsahuje širokou škálu změřených nukleových kyselin mnoha organismů. V této kapitole se zaměřím na datové formáty, které lze využít ke zpracování informací z tandemových hmotnostních spektrometrů a na formát t2d, který slouží k zápisu dat získaných z přístroje 4700 MALDI TOF/TOF (Applied Biosystems 4700 Proteomics Analyzer).

### 4.1 mzML formát ve vztahu k MS/MS

Zavedený standard uložení naměřených „mz“ dat (mzXML) se ukazuje jako příliš těžkopádný. Pracovní skupina pro standardy hmotnostní spektrometrie se zabývá definicemi datových formátů a archivací v hmotnostní spektrometrii proteomu.

V letech 2005-2008 existovaly dva oddělené XML formáty pro kódování surových výstupních dat ze spektrometru:

- mzData vyvinutý PSI
- mzXML vyvinutý v Ústavu pro systémové biologie (ISB).

Je známo, že tyto dva odlišné formáty jsou v podstatě totéž. Proto PSI, s plnou účastí ISB, vyvinul nový formát s nejlepšími aspekty z každého z předchozích formátů. Ta měla nahradit předchozí dva formáty. Tento nový formát byl původně uveden pracovním názvem dataXML [23].

### 4.2 mzXML formátu ve vztahu k MS/MS

Data získaná z tandemového hmotnostního spektrometru jsou uložena v jednom z nejpoužívanějších formátů *mzXML*.

*mzXML* je otevřený datový formát pro ukládání a výměnu dat s hmotnostní detekcí, vyvinutá v *SPC (Institut pro systémové biologie)*. *mzXML* poskytuje standardní zásobník pro MS a MS/MS proteomové údaje a je základem dalších proteomových sítí. Proprietární formáty souborů od většiny výrobců mohou být přeměněny na otevřený *mzXML* formát. Původním autorem tohoto formátu je Patrick Pedrioli. V současné době existuje několik verzí tohoto formátu. Jedná se o 1.0 (také zvanou „MSXML“), 2.0, 2.1, 3.0 a 3.1 (aktuální verze)[24].

### 4.3 t2d formát ve vztahu k MS

Formát používaný společností Applied Biosystems. Tento formát dat je obvykle získáván ze 4700 a 4800 MALDI TOF/TOF hmotnostních spektrometrů. Data, která získáme, jsou v binární podobě. To znamená, že nemají v textovém režimu čitelnou formu a pro jejich čtení je potřeba speciální program [32].

### 4.4 Proteom

Aplikace pro převod různých formátů dat, která jsou získána z různých hmotnostních spektrometrů na jiné formáty, které jsou „čitelné“. Je vytvořený v aplikační platformě JAVA a díky tomu lze použít její knihovny v programu Matlab. Po načtení knihovny Proteomu do Matlabu je zde možné číst formát t2d přímo. Proteom poskytuje sadu funkcí pro práci s formátem t2d a vrací pole, které je zpracovatelné v Matlabu a obsahuje uspořádané párové dvojice intenzita (osa y) a hodnota m/z (osa x). Tato data jsou již snadno zpracovatelná v Matlabu pomocí bioinformatického toolboxu [33].

## 5 Flavonoidní sloučeniny

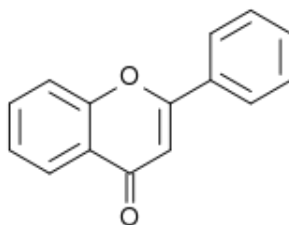
Flavonoidy patří mezi velmi bohatou skupinu sekundárních rostlinných metabolitů s fenylobenzopyronovým jádrem, které se řadí mezi sekundární metabolity produkované rostlinnými druhy. V současné době je známo více než 5000 druhů flavonoidních látek. Tyto látky se nezúčastňují přímo primárních metabolických procesů (fotosyntéza, respirace), ale pomáhají rostlinám reagovat na změny podmínek životního prostředí nebo na ataky různých patogenů.

Význam flavonoidů pro člověka je velmi důležitý. Jsou především významnou součástí antioxidačního systému, zabráňují peroxidaci lipidů, likvidují volné radikály, mohou vázat a inaktivovat některé prooxidační kovové ionty (Cu, Fe), jsou významné v prevenci proti nádorovým onemocněním, kardiovaskulárním chorobám a klimakterickým obtížím. Tyto látky byly prokázány v mnoha potravinách, které jsou pro nás zcela běžné (čokoláda, pomeranče, grepfruity). Účinně také flavonoidy působí proti poškození cév a mají rovněž vliv na aktivitu enzymů [34].

Studium působení těchto přírodních produktů nabízí i mnohá praktická využití. Například pozorováním přirozených obranných mechanismů rostlin proti infekcím a predátorům, aby se dalo zredukovat užívání drahých a potencionálně toxických pesticidů. Poznatky z oblasti genového inženýrství se dají využít ve farmacii, parfumerii a při výrobě materiálů pro komerční využití [35].

### 5.1 Chemická stavba flavonoidů

Základ struktury flavonoidů tvoří flavan, který se skládá ze dvou benzenových jader, které jsou spojeny pyranem. Hydroxyskupiny a keto skupiny, substituované na tuto základní strukturu, odlišují jednotlivé skupiny flavonoidů. Přírodní flavonoidy zaujímají nejčastěji podobu O-glykosidů, jejich molekula je tedy tvořena cukernou částí a částí necukernou (aglykonem) [36].

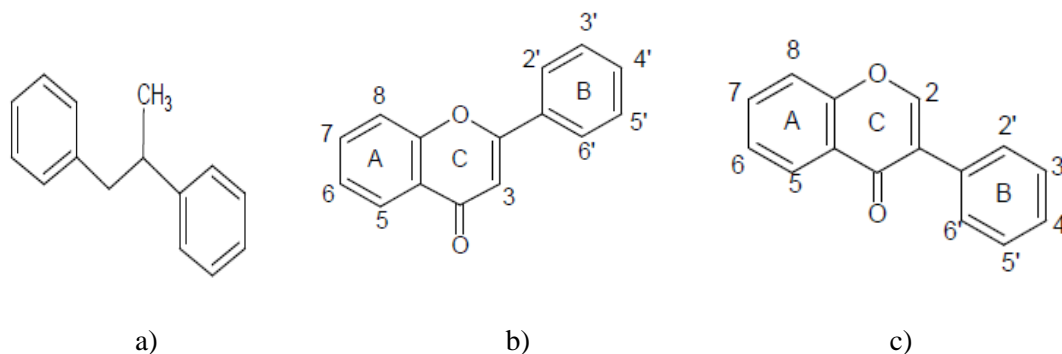


**Obrázek 21:** Základní struktura flavonoidů [36]

### 5.2 Isoflavonoidy

Isoflavonoidy tvoří rozsáhlou podskupinu polyfenolických biologicky aktivních přírodních flavonoidních sloučenin. Patnáctiuhlíkový skelet isoflavonoidů ( $C_6-C_3-C_6$ ), je formálně odvozen od 1,2-diphenylpropanového skeletu.

V rostlinách jsou isoflavonoidy obsaženy jako volné aglykony, méně často jako glykosidové deriváty (obecně jako konjugáty s cukernou složkou glukózou, rhamnosou či apiosou). Základní skelet molekul isoflavonoidů (3-fenylchromen-4-on) podléhá substitucím (methylace, prenylace, hydroxylace, chlorace, připojení aromatických či alifatických kyselin, aminoskupin apod.) a cyklizacím. Isoflavonoidy se vyskytují v různých oxidačních stupních a dále ve formě dimerů a heterodimerů [35].



**Obrázek 22:** Difenylpropanový skelet - 1,2-difenylpropan (a), základní skelet flavonu - flavon (2-fenylchromen-4-on) (b) a isoflavonu - isoflavon (3-fenylchromen-4-on) (c) [36]

### 5.2.1 Dělení isoflavonoidů do podskupin

Isoflavonoidy jsou na základě své strukturní rozmanitosti děleny do následujících podskupin:

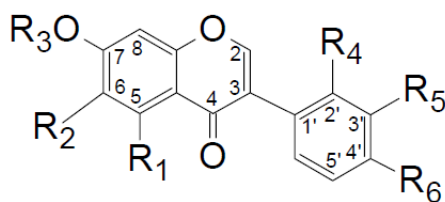
- Isoflavony
- Isoflavany
- Isoflavanchinony
- Isoflavanony
- Isoflav-3-eny
- Rotenoiny

Nejpočetnější skupinu isoflavonoidů tvoří isoflavony.

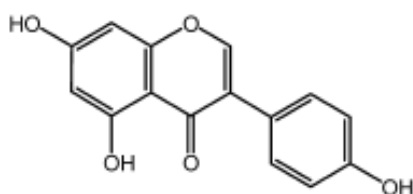
### 5.2.2 Isoflavony

Isoflavony patří (Obrázek 23) společně s flavanony, flavonoly, katechininy, anthocyanidininy a dalšími látkami) mezi velmi rozšířenou skupinu přírodních produktů s názvem flavonoidy. Tvoří sekundární metabolity rostlin a dokážou vyvolat obrovský počet fyziologických reakcí nejen u člověka, ale i u dalších savců. Isoflavony jsou dobře známé jako protinádorové látky a našli také využití v prevenci kardiovaskulárních nemocí. Kromě kladných účinků isoflavonů na lidské zdraví, existuje i několik publikací o jejich toxických účincích, které jsou spojeny s množstvím aplikované dávky a dalšími faktory. Nejznámějšími a nejprozkoumanějšími isoflavony jsou genistein (Obrázek 24) a daidzein (Obrázek 25) [37].

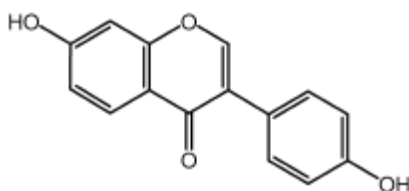
Přehled struktury isoflavonů, které mohou být navaázány na jednotlivé konce základní struktury isoflavonů (Obrázek 23) je uveden v příloze (Příloha 2)



**Obrázek 23:** Základní struktura isoflavonů [37]



**Obrázek 24:** Genistein [38]



**Obrázek 25:** Daidzein [39]

### 5.2.3 Výskyt isoflavonů

V současné době je známo více než 300 rostlin, produkujících fytoestrogeny (Reynaud et al., 2005). Patrně nejvýznamnějším producentem isoflavonoidních fytoestrogenů je sója (*Glycine max*) z čeledi bobovitých. Ta obsahuje 50-300 mg především genisteinu a daidzeinu na 100 g plodů, dále Vigna sp., cizrna (*Cicer sp.*), vojtěška (*Medicago sativa*) a jetel (*Trifolium sp.*).

Mimo tuto čeleď byly isoflavonoidy detekovány v kosatci (*Iris sp.*), semenech sezamu (*Sesamum indicum*), slunečnice (*Helianthus sp.*) a máku (*Papaver sp.*), v červeném zelí, brokolici, květáku (vše *Brassica oleracea spp.*) a dalších plodinách. Z obilí ječmene (*Hordeum vulgare*) či samičích šištic chmele (*Humulus lupulus*) pocházejí isoflavony daidzein, genistein, které byly v nízkých koncentracích objeveny v pivu [36].

### 5.2.4 Isoflavony v rostlinách čeledi bobovité

#### Zařazení bobovitých do rostlinného systému:

Nadříše: Jaderní (*Eukaryota*)

Říše: Rostliny (*Plantae*)

Podříše: Vyšší rostliny (*Eumobionta*)

Oddělení: Krytosemenné (*Magnoliophyta*)

Třída: Dvouděložné (*Magnoliopsida*)

Čeleď: Bobovité (*Fabaceae*) [40]

### Rod: Jetel (*Trifolium*)

Rod jetelovitých rostlin (*Trifolium*) obsahuje celkově asi 300 druhů, v ČR se pak můžeme setkat s 23 druhy. Rozšíření těchto rostlin je velmi kosmopolitní, ale největší výskyt je v mírném pásu severní polokoule. Jetelovité rostliny jsou nízké jednoleté, dvouleté nebo vytrvalé byliny. Listy jsou trojčetné, opatřené palisty na spodní části řapíku. Lístky jsou přisedlé, prostřední někdy s krátkým řapíkem. Drobné květy jsou zbarvené červeně, růžově, bíle nebo žlutě, skládají se do hlávek. Plody jsou drobné lusky uzavřené v kalichu. Některé druhy poskytují velice kvalitní píci, proto jsou také často pěstovány na zemědělských plochách. Díky účinné mykorhize dokážou efektivně vázat dusík, proto jsou vhodné i pro zkvalitnění půdy.

Mezi zástupce tohoto rodu patří druhy: jetel horský (*Trifolium montanum*), jetel luční (*Trifolium pratense*), jetel plazivý (*Trifolium repens*), jetel prostřední (*Trifolium medium*) a jetel zvrhlý (*Trifolium hybridum*) [41].



Obrázek 26: Jetel luční (*Trifolium pratense*) [42]

### Výskyt isoflavonů v jetelovitých rostlinách

Hlavním zdrojem isoflavonů v Asii je sója. Pokusy o konzumaci většího množství sójového mléka v západní civilizaci ztroskotaly díky vedlejšímu efektu sóji - navyšování hmotnosti. Západní civilizace má však další handicap: častá konzumace alkoholu, „vymazává“ příznivé účinky isoflavonů. Zcela výjimečným zdrojem isoflavonů je jetel luční nebo také jetel červený (*Trifolium pratense* L.). Tato rostlina používaná jako léčivo v Evropě i v Orientu proti ekzému a psoriase, má i účinky spasmolytické a diuretické. Obsahuje všechny klinicky účinné isoflavony.

Ty jsou extrahovány z jeho lístků, přičemž květy obsahují isoflavonů jen velmi málo. Jetel červený je pěstován za přísných vegetačních podmínek a není geneticky modifikován. Má 10-20x vyšší hladinu isoflavonů než sója, která je druhým nejbohatším zdrojem. Jetel červený je dále jednou z mála rostlin, která obsahuje v extraktu všechny farmakologicky aktivní isoflavony. Navíc, a to se

dnes jeví jako jedna z hlavních výhod červeného jetele, je to pro nás, středoevropany, „tradiční“ rostlina a náš organismus je mnohem lépe připraven na její „zpracování“, než je tomu například u sóji. Isoflavony jsou totiž v sóji vázány na bílkoviny, což snižuje jejich využitelnost v našem organismu a nevýznamné není ani společenské hledisko, tedy že při trávení sóji dochází u mnoha jedinců k meteorismu [43].

### **Rod: Sója (*Glycine*)**

Rod Sója (*Glycine*) patří mezi luštěniny (společně s hrachem, fazolí a čočkou) a zahrnuje několik druhů. Nejznámější a nejrozšířenější z nich je pak druh sója luštinatá (*Glycine max*), která je považována za jednu z nejstarších kulturních rostlin, původem z jižní Asie a její pěstování je rozšířeno prakticky po celém světě.

Sója luštinatá je jednoletá, v případě některých kultivarů dvouletá rostlina, která dosahuje výšky od 0,5 do 1m, rezavě hnědé barvy, chlupatá s přímou, větvenou lodyhou s listovým, podobným listoví fazolovému. Listy jsou trojčetné, palistnaté, lístky okrouhle vejčité až eliptické, celokrajné s drobnými palísky. Květní hrozny jsou úžlabní a mají 5 až 8 květů, kultivary 15 až 20 květů. Kalich je trubkovitý, koruna starorůžová, lehce fialová nebo i bílá. Plodem jsou lusky 3 až 8 x 1cm, dvou až čtyřsemenné. Kvete v červenci a srpnu.



**Obrázek 27:** Sója luštinatá (*Glycine max*) [42]

Vliv sóji na lidský organismus je velmi prospěšný, blahodárně působí především na tělní oběh, krevetvorbu, nervovou soustavu a látkovou výměnu. Při pravidelné konzumaci snižuje hladinu cholesterolu v krvi, upravuje krevní tlak, chrání před tvorbou žlučových kamenů a snižuje riziko rakoviny prsu. Je zdrojem plnohodnotných bílkovin, minerálních látek (draslík, hořčík, vápník, fosfor, měď atd.). Při nadměrné konzumaci má však negativní vliv na životnost červených krvinek.

Isoflavony byly v sóji nalezeny především v chloroplastech nadzemních částí rostlin, ve stopovém množství i v kořenech. Vyskytují se jako látky konstituční nebo se objevují jako výsledek působení stresu. Isoflavony plní určité funkce v obranném systému rostliny jako přirozená ochrana proti infekci při klíčení semen, napadení hmyzem a poškození škůdci [44].

**Tabulka 3:** Množství isoflavonů v různých částech rostliny sóji luštinaté (*Glycine max*) [44]

	Zastoupení isoflavonů v µg/g (vážený průměr)					
Isoflavony	Rostlina	Kořen	Lodyha	List	Lusk	Semeno
Daidzin	7,71	9,53	4,75	7,05	2,12	11,52
Genistin	9,29	3,79	2,94	4,57	1,28	12,18
Ononin	1,96	4,14	1,43	2,22	-	-
Daidzein	1,47	7,38	0,91	-	1,18	0,98
Genistein	0,88	1,34	0,8	1,01	0,74	0,89
Formononetin	0,83	1,46	-	0,8	0,89	-

### 5.2.5 Isoflavony v rostlinách čeledi konopovité

#### Zařazení konopovitých do rostlinného systému:

Nadříše: Jaderní (*Eukaryota*)

Říše: Rostliny (*Plantae*)

Podříše: Vyšší rostliny (*Cormobionta*)

Oddělení: Krytosemenné (*Magnoliophyta*)

Třída: Dvouděložné (*Magnoliopsida*)

Čeleď: Konopovité (*Cannabaceae*) [40]

#### Rod: Chmel (*Humulus*)

Chmel zahrnuje 2 druhy, které jsou původní v mírném pásmu, v jiných oblastech světa jsou pěstovány a místy i zplaňují. V ČR roztroušeně zejména v teplejších oblastech (max. asi 700 m n.m.). Celkově je původní v téměř celém mírném pásmu Eurasie, pěstováním byl však zavlečen i do Severní a Jižní Ameriky, Austrálie a dalších částí světa.

Pravotočivá oplétavá dvoudomá bylina s lodyhami až 6m dlouhými, s listy dlanitě 3 -5 laločnými, na okraji pilovitými. Listy mají na bázi řapíku dva blanité palisty. Prašníková květenství v úžlabí nebo konečná jsou sestavena v latách, pestíková se skládají z krátkých klásků a vytvářejí se z nich vejčité šištice. Jejich šupiny až 2 cm dlouhé jsou pokryté zlatožlutými lupinovitými žlázkami. Pro lékařské a průmyslové účely mají význam jen květenství samičí, pestíková. Šištice chmele rostoucího planě nebo zplaněle jsou pro léčebné účely bezcenné. Kvete od července do srpna.





**Obrázek 28:** Chmel (*Humulus lupulus*)[45]

Sbírají se samičí šištice nebo chmelové (lupulinové) žláзки zvané lupulin, které vyrůstají na listenech tvořících šištici. Šištice se sbírají ve druhé polovině srpna nebo začátkem září, kdy obsahují nejvíce účinných látek. Suší se za teplot nepřesahujících 40 °C. Lupulin se získává proséváním a dalším čištěním chmelových šištic.

Obsahuje pryskyřice, hořčiny (např. humulon, lupulon a další), silice s terpenoidy (např. humulen, myrcen aj.), flavonoidy, fytoncidy, fytosteroly, vosky, cholin aj. Chmel působí sedativně, podporuje trávení a má desinfekční účinky. Užívá se při nespavosti, nervovém rozrušení, neklidu, nadýmání nebo při potížích spojených s klimakteriem.

Příznivé působení chmelu (piva) na srdce a cévní systém lidského organismu je velmi významný. Podstatný je komplexně příznivý význam, na němž se podílí řada látek obsažených v pivu a to různými mechanismy a na různých úrovních. Zásadní ochranný význam mají ve chmelu obsažené polyfenoly, flavonoidy a vitamíny skupiny B při vzniku a rozvoji aterosklerózy s následky v podobě ischemické choroby srdeční, infarktu myokardu, cévní mozkové příhody – mrtvice, ischemické choroby dolních končetin apod. Významné je i působení flavonoidů na funkci krevních destiček, které brání vzniku krevních sraženin. Polyfenoly a flavonoidy zasahují do oxidačně-redukčních pochodů v organismu a mají antioxidační účinky, které vedou k tvorbě chelátových vazeb s kovovými ionty, a proto působí jako „zametači“ volných radikálů [45].

### 5.3 Izolace a identifikace isoflavonů

Pro izolaci isoflavonů se používá rostlinný materiál čerstvý, sušený nebo lyofilizovaný, přičemž velmi bohatým zdrojem aglykonů jsou oddenky, kořeny, dřevo a kůra, méně pak listy a květy rostlin. Obsah isoflavonů v rostlině je však závislý na velké řadě faktorů: půdní a klimatické podmínky, druh a stáří rostliny, přítomnosti patogenů a symbiontů atd.

Extrakce se provádí až čtrnáctidenní macerací za laboratorní teploty rozpouštěním ve vodných roztocích ethanolu, metanolu nebo acetonu. Získaný extrakt se dále zbavuje hydrofobních látek rozpouštěním v hexanu či jiném nepolárním rozpouštědle a následně zpracovává moderními technikami.

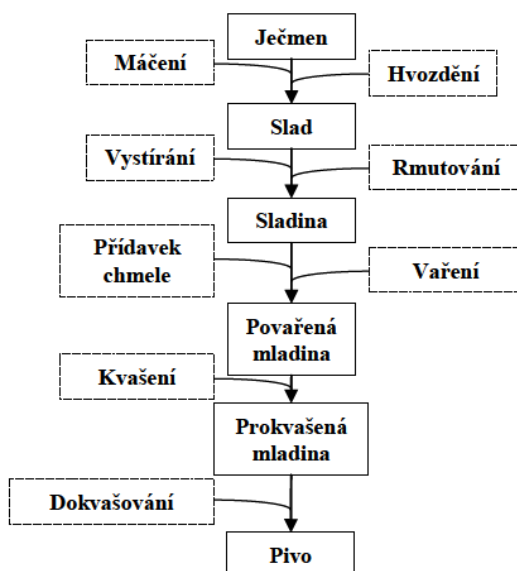
V současnosti je nejběžnější metodou pro identifikaci isoflavonů vysokoúčinná kapalinová chromatografie (HPLC) a hmotnostní spektrometrie (MS), často spojená s chemickou ionizací za atmosférického tlaku (APCI). Pro identifikaci komplexních molekul je vhodná dvourozměrná nukleární magnetická rezonance (2D-NMR). Účinným nástrojem pro analýzu isoflavonů je rovněž

kapilární elektroforéza (CE) kombinovaná s ionizací elektrosprejem a hmotnostní spektrometrií (ESI-MS) [36].

## 5.4 Sladina a mladina ve výrobě piva

Základními látkami pro výrobu piva jsou ječmen a chmel (Obrázek 29). Pivo je připravováno fermentací ze sladovaných obilnin, především z ječmene. Při sladování dochází nejprve v obilkách ječmene k cílenému zvýšení obsahu vody (metoda tzv. máčení). Tím dochází k zahájení procesů, které jsou propojené s klíčením, především enzymatické reakce. Ve vhodném okamžiku se zvýšením teploty (tzv. hvozdění) zastaví metabolické pochody klíčků a také většina enzymatických dějů.

Po rozemletí sladu jsou látky v něm obsažené převedeny do roztoku. Proces je nazván vystírání. Aktivita enzymů, které nebyly denaturovány při hvozdění, je teplotně řízena (při tzv. rmutování), aby bylo dosaženo optimálního složení vznikající sladiny. Následným povařením s chmelem se nakonec ukončí i tyto procesy a vznikne povařená mladina. Procesem zvaným fermentace vzniká pomocí kvasinek prokvašená mladina, která je velmi podobná výslednému produktu [49].



**Obrázek 29:** Zjednodušené schéma výroby piva [49]

## 6 Realizace aplikace

Praktická realizace mé bakalářské práce spočívá v analýze získaných spekter, která byla naměřena pomocí hmotnostního spektrometru Applied Biosystems 4700 Proteomics Analyzer (Applied Biosystems, Framingham, USA), který pracuje na principu metody MALDI-TOF. Analýza je provedena pomocí aplikace vytvořené v grafickém uživatelském prostředí (GUI) Matlab (verze 2008b). Vývojový program aplikace v Matlabu je zobrazen v příloze (Příloha 3).

### 6.1 Požadavky na aplikaci pro analýzu

Požadavky na vytvořenou aplikaci byly následující:

1. Získání spekter ze souboru, která jsou ve formátu t2d.
2. Zobrazení těchto spekter a aplikace možnosti ořezání a prahu vyhledání píků.
3. Zobrazení teplotní mapy (*heat map*) těchto spekter.
4. Filtrace pomocí VF filtru, rozpoznání a zobrazení píků ve vyfiltrovaném spektru (*peak recognize*).
5. Sloučení dvou načtených spekter pomocí aritmetického průměru a jejich následné vykreslení.
6. Spojení dvou načtených spekter pomocí jejich rozdílu a následné vykreslení.
7. Pro obě dvě spojení vypočítání a zobrazení statistických parametrů (korelační koeficient, rozptyl, střední hodnota).
8. Zobrazení korelační křivky vztahu dvou načtených spekter.
9. Uložení zpracovaných výsledků do souboru formátu typu csv.

### 6.2 Hmotnostní spektrometrie v bioinformatickém toolboxu programu Matlab

Aplikace je vytvořena pomocí bioinformatického toolboxu programu Matlab. Tento toolbox je určen pro zpracování bioinformatických dat a z jeho možných funkcí jsem použila funkce pro zpracování MS spekter.

Přehled funkcí a jejich vlastností jsou shrnuty v následující tabulce [48].

**Tabulka 4:** Přehled funkcí Bioinformatického toolboxu pro zpracování MS spekter

Funkce	Význam
isotopicdist	Výpočet izotopu s vysokým rozlišením hmotnostního rozložení a hustoty
jcampread	Přečte soubory ve formátu JCAMP-DX
msalign	Zarovnání píků v signálu na referenční píky
msbackadj	Určí správnou základní linii signálu s píky
msdotplot	Vykreslí píky z datového zdroje LC/MS nebo GC/MS
msheatmap	Vytvoří pseudobarevný obrázek teplotní mapy z hmotnostního spektra
mslowess	Vrací vyhlazený signál s vrcholy pomocí neparametrické metody
msnorm	Normalizuje signál s píky
mssalign	Zarovná hmotnostní spektrum z píky datových zdrojů LC/MS nebo GC/MS
mspeaks	Vyhledá důležité píky ve spektru podle zadaných parametrů
msspresample	Opětovné navzorkování signálu s píky při zachování původních píků
msresample	Znovu navzorkování signálu s píky
mssgolay	Získá polynomicky hladký signál s píky metodou nejmenších čtverců.
msviewer	Zobrazí hmotnostní spektrum, nebo množinu hmotnostních spekter
mzcdf2peaks	Převede mzCDF strukturu do seznamu píků
mzcdfinfo	Vrátí informace o souboru mzCDF, který obsahuje data hmotnostních spekter
mzcdfread	Přečte mzCDF soubor a vrátí hmotnostní spektrum
mzxml2peaks	Převede mzXML strukturu do seznamu píků
mzxmlinfo	Vrátí informace o souboru mzXML, který obsahuje data hmotnostních spekter
mzxmlread	Přečte mzXML soubor a vrátí hmotnostní spektrum
samplealign	Zarovná dvě množiny dat, které obsahují sekvenční pozorování pomocí mezer
tgspcinfo	Vrátí informace o souboru SPC, který obsahuje data hmotnostních spekter
tgspcread	Přečte SPC soubor a vrátí hmotnostní spektrum

## 6.3 Popis implementace

### 6.3.1 Způsob načtení t2d dat

K načtení formátu t2d jsem použila Java knihoven programu Proteom. Tyto knihovny jsem dále načetla do aplikace v Matlabu.

Knihovny v programu Proteom poskytují sadu funkcí pro zpracování t2d formátu a vrací strukturu, která je použitelná k zpracování v Matlabu. Funkce, kterou jsem k tomuto kroku použila, se jmenuje *T2DPeakListReader*. Jako vstupní parametr přijímá cestu k souboru, který je ve formátu t2d.

```

t2dFile = org.proteomecommons.io.t2d.T2DPeakListReader([PathName
FileName]);
peakList = t2dFile.getPeakList;
peaks = peakList.getPeaks;
ind = numel(peaks);
MZ = zeros(ind,1);
Y = zeros(ind,1);
for count = 1:ind
    MZ(count) = peaks(count).getMassOverCharge;
    Y(count) = peaks(count).getIntensity;

```

end

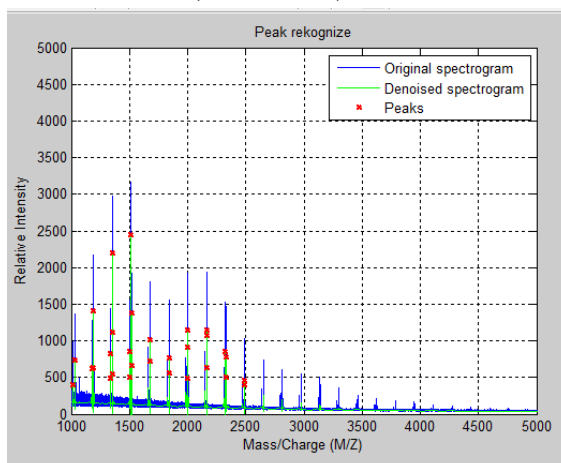
```
spectr1.data(:,1) = MZ;  
spectr1.data(:,2) = Y;
```

Ve výše uvedené části programovacího kódu je vidět, že program nejprve získá pole objektu typu *peaks*, ze kterého jsou následně v cyklu získány dvě pole obsahující poměr hmotnosti a náboje (*MZ*) a intenzitu pro každý poměr hmotnosti a náboje (*Y*). Tyto pole jsou pak nahrány do jedné proměnné a tato proměnná je uložena na disk pro další zpracování.

### 6.3.2 Zobrazení teplotních map (heat map) a rozpoznání píků (peak recognize)

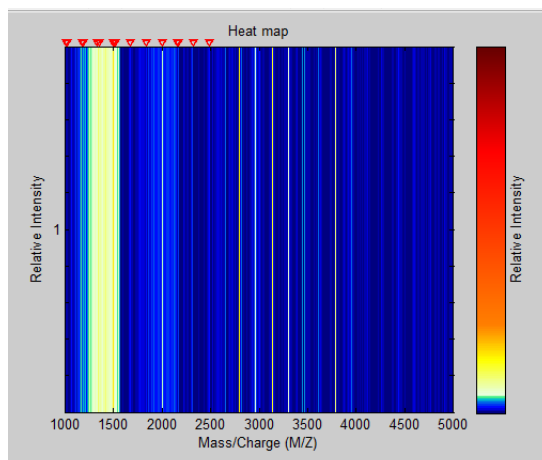
K implementaci těchto funkcí byly použity funkce z bioinformatického toolboxu (*Bioinformatic toolbox*), konkrétně:

- *Mspeaks* – tato funkce filtruje zpracovávaná spektra a vrací pole, které obsahuje rozpoznané píky. Zároveň zobrazuje i graf, ve kterém je vykresleno původní spektrum, vyfiltrované spektrum a vyznačené píky ve filtrovaném spektru. Filtrace je provedena pomocí vysokofrekvenčního filtru (Obrázek 30).



Obrázek 30: Zobrazení píků pomocí funkce *Mspeaks*

- *Msheatmap* – pomocí tohoto příkazu zobrazím teplotní mapu zkoumaného spektra s vyznačenými píky. Nejintenzivnější pík zde má barvu jasně červenou, nejméně intenzivní pak barvu modrou. Tato funkce nám umožňuje lepší a přehlednější orientaci ve spektru (Obrázek 31).



Obrázek 31: Zobrazení píků pomocí funkce *Msheatmap*

Obě tyto funkce přijímají jako vstupní parametry pole  $MZ$  a  $Y$ . Tato pole jsou základně nastavena pro zobrazení v rozmezí hodnot  $MZ$  od 1000 do 5000, ale mohou být „ořezána“ podle parametrů zadaných uživatelem.

Mez intenzity, od které je pík považován za „použitelný“ je nastavena na hodnotu 400, ale tato hodnota může být opět změněna uživatelem.

### 6.3.3 Získání sloučených spekter

Spektra jsou slučována pomocí dvou principů a to konkrétně aritmetickým průměrem intenzity obou spekter a rozdílem intenzity spekter. Tento úkon je proveden jednoduchými funkcemi v Matlabu.

Takto získaná nová spektra se opět zobrazí a jako u původních načtených spekter, lze i těchto vykreslit teplotní mapu (*heat map*) a rozpoznání píků (*peak racognoize*).

Tato spektra jsou použita pro následnou analýzu, a proto jsou brány jako výsledek a cíl této aplikace.

Dále se k těmto spektrům zobrazí statistické parametry, pro jejichž výpočet jsou použity funkce ze statistického toolboxu (*Statistics Toolbox*):

- `Corrcoef` – korelační koeficient. Tato funkce vrací korelační koeficient.
- `Mean` – střední hodnota. Funkce slouží pro výpočet střední hodnoty získaných spekter.
- `Std` – rozptyl. Rozptyl získaných spekter je vypočítán touto funkcí.

Hodnoty získané pomocí statistických parametrů jsou vypsány do dvou oddělených tabulek. Každé sloučené spektrum má vlastní.

Jako další informační hodnota je zobrazena korelační křivka obou původních spekter. K vytvoření křivky byla použita funkce `xcorr`.

### 6.3.4 Uložení získaných dat do souboru csv

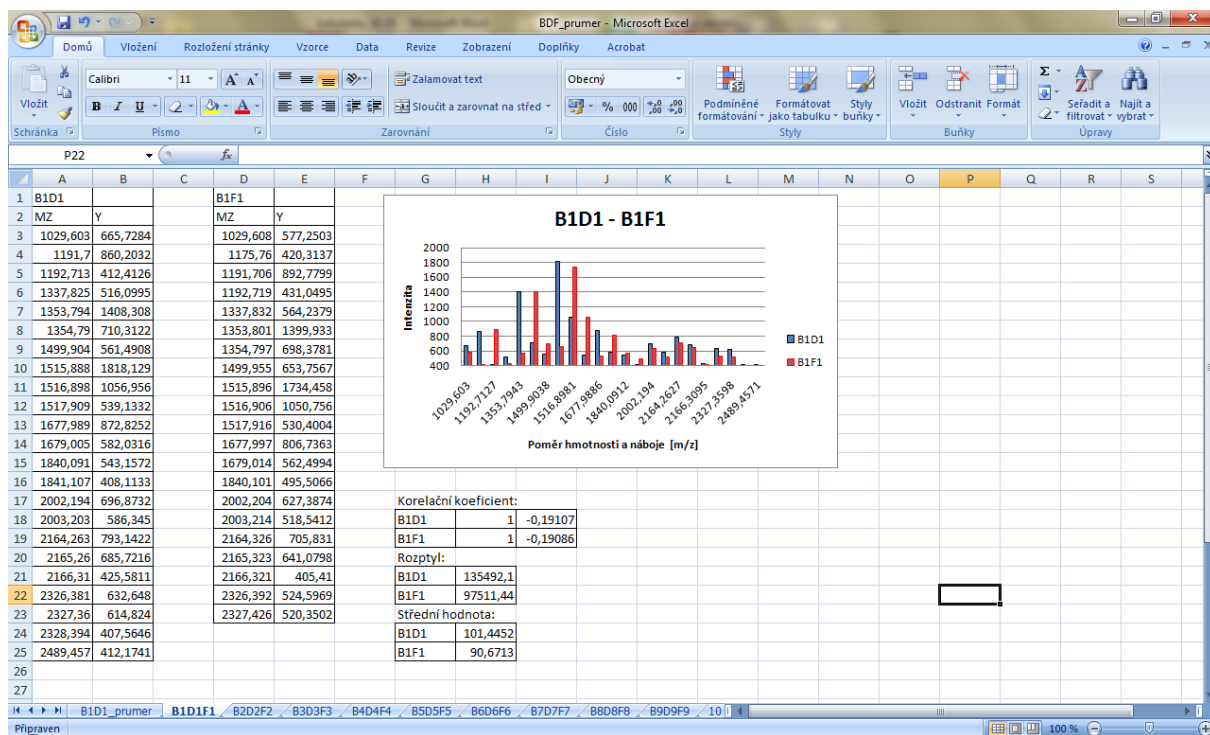
Rozpoznané píky mohou být pro další zpracování a analýzu ukládány do souboru formátu csv. Tento formát umožňuje načtení zpracovaných dat v programu MS Excel (Obrázek 32). Uživatel musí zadat název souboru do textového pole v GUI. Analyzovaná data jsou poté uložena do dvou souborů, kde první obsahuje data zpracována aritmetickým průměrem a druhý rozdílem. Do tohoto souboru jsou data uložena ve dvou sloupcích a to  $MZ$  (poměr hmotnosti a náboje) a  $Y$  (intenzita).

### 6.3.5 Uložení získaných statistických parametrů

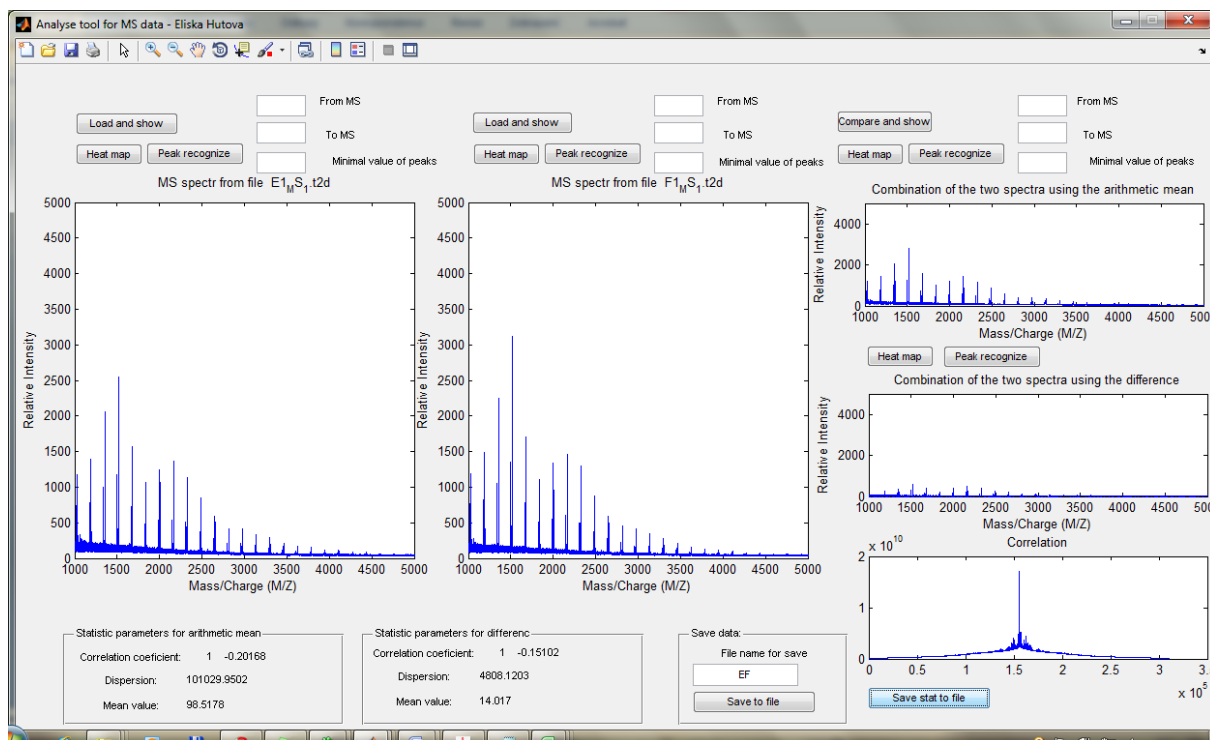
Získané statistické údaje jsou uloženy do souboru formátu mat. Pro snadnější orientaci uživatele jsou uloženy ve dvojicích:

- *název* – hodnota zadaná uživatelem do pole názvu souboru;
- *hodnota* – daný statistický parametr.

Na požádání uživatele mohou být všechna data z provedené analýzy uložena do souboru typu csv.



Obrázek 32: Zobrazení csv dat v programu MS Excel



Obrázek 33: Vzhled GUI aplikace

## 7 Popis aplikace a zpracování výsledků

### 7.1 Spektra testovacích a modelových isoflavonů

Jako testovací spektra byly využity vzorky spekter isoflavonů chmele. Jednotlivé skupiny souborů byly označeny písmeny A, B, C, D, E, F, kde A a B jsou spektra ze sladiny (první skupina), C a D jsou spektra povařeně mladiny (skupina druhá) a E a F jsou spektra z prokvašené mladiny (skupina třetí).

Modelově jsem porovnávala porovnání skupin spekter, kdy jsem po zpracování spekter v programu Matlab získala aritmetický průměr spekter AB, CD a EF. Tato data jsem dále vyhodnotila v programu MS Excel. Data jsou přehledně rozložena v jednotlivých tabulkách a ke každé dvojici tabulek je vytvořen graf. Grafy jsou sloupcového typu, kde výsledná spektra jsou zobrazena na ose  $x$  (poměr hmotnosti a náboje  $m/z$ ) od hodnoty 1000 do hodnoty 2500 a na ose  $y$  (intenzita) jsou vykreslena od hodnoty 400 (mez intenzity), protože v nižších hodnotách jsou spektra příliš zatížena šumem.

Pro lepší vyhodnocení naměřených a zpracovaných spekter jsou použity statistické parametry.

### 7.2 Statistické parametry v analýze spekter

Při vyhodnocení zpracovaných spekter jsem použila statistické parametry, které upřesní analýzu získaných spekter [46].

#### 7.2.1 Rozptyl $D(x)$

Rozptyl je mírou variability náhodné veličiny.

Rozptyl náhodné veličiny s diskrétním rozdělením s pravděpodobnostní funkcí  $P(x)$  je definován vztahem [47]:

$$D(X) = \sum_r (x - E(X))^2 P(x). \quad (15)$$

Rozptyl lze vypočítat podle vzorce:

$$D(X) = E(X - E(X))^2 = E(X^2 - 2XE(X) + (E(X))^2) = [E(X)]^2 \quad (16)$$

#### 7.2.2 Střední hodnota $E(x)$

Střední hodnota náhodné veličiny  $X$  s diskrétním rozdělením daným pravděpodobnostní funkcí  $P(x)$  je definována vztahem [47]:

$$E(X) = \sum_x xP(x) \quad (17)$$

#### 7.2.3 Korelační koeficient $\rho_{xy}$

Dává určitou informaci o stupni závislosti dvou náhodných veličin. Je definován jako poměr kovariance k součinu směrodatných odchylek obou náhodných veličin.

Korelační koeficient  $\rho_{xy}$  dvou náhodných veličin  $X$  a  $Y$  s rozptyly  $\sigma_x^2 > 0$  a  $\sigma_y^2 > 0$  je definován vztahem:



$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \quad (18)$$

Je-li  $\sigma_x^2 = 0$  nebo  $\sigma_y^2 = 0$  pokládáme  $\rho_{xy} = 0$  [47].

### 7.3 Vyhodnocení metody porovnání spekter metodou aritmetického průměru

Při porovnání spekter jednotlivých skupin (AB, CD a EF) jsem po provedené analýze v programu Matlab, uložení dat do souboru a provedeném vyhodnocení v MS Excel byly získány údaje, z nichž jsem vyvodila následující závěry:

1. Ze získaných statistických parametrů (Tabulka 5 - 5) lze usoudit, že mladina (povařená i prokvašená) je ve svých statistických hodnotách velmi podobná oproti sladince (skupina AB).
2. Spektra skupiny AB (sladina) obsahují méně rozpoznávaných píků než spektra skupiny CD a EF (povařená a prokvašená mladina).
3. Spektra skupiny AB nemají tak velké rozpětí hodnot na ose  $x$  (poměr hmotnosti a náboje  $m/z$ ). V tabulce (Tabulka 8) je názorně zobrazeno, že spektra skupiny AB jsou na ose  $x$  vyhodnoceny pouze do hodnoty 2327,28, kdežto spektra skupiny CD (resp. EF) jsou vykresleny až do hodnoty 2490,47 (2489,526).
4. Výsledná intenzita píků. V grafu (Obrázek 34) je názorně vidět, že nejintenzivnější pík skupiny AB dosahuje hodnoty intenzity 1446,54 Da, zatímco nejintenzivnější (nejvyšší) píky skupin CD a EF dosahují hodnot intenzity 2215,25 Da a 2302,78 Da.

Z výše uvedených bodů lze usoudit, že sladina obsahuje obecně méně flavonoidních sloučenin (isoflavonů) než mladina jak povařená, tak i prokvašená. Z grafu lze dále usoudit, že prokvašená mladina je ale oproti mladince povařené tvořena větším množstvím isoflavonů. Tento stav je způsoben výrobními procesy zkoumaných látek.

**Tabulka 5:** Korelační koeficient spekter skupin AB, CD, EF

Korelační koeficient					
AB		CD		EF	
1	-0,16723	1	-0,19625	1	-0,20168

**Tabulka 6:** Rozptyl spekter skupin AB, CD, EF

Rozptyl		
AB	CD	EF
84889,91	118979,4	101030

**Tabulka 7:** Střední hodnota spekter skupin AB, CD, EF

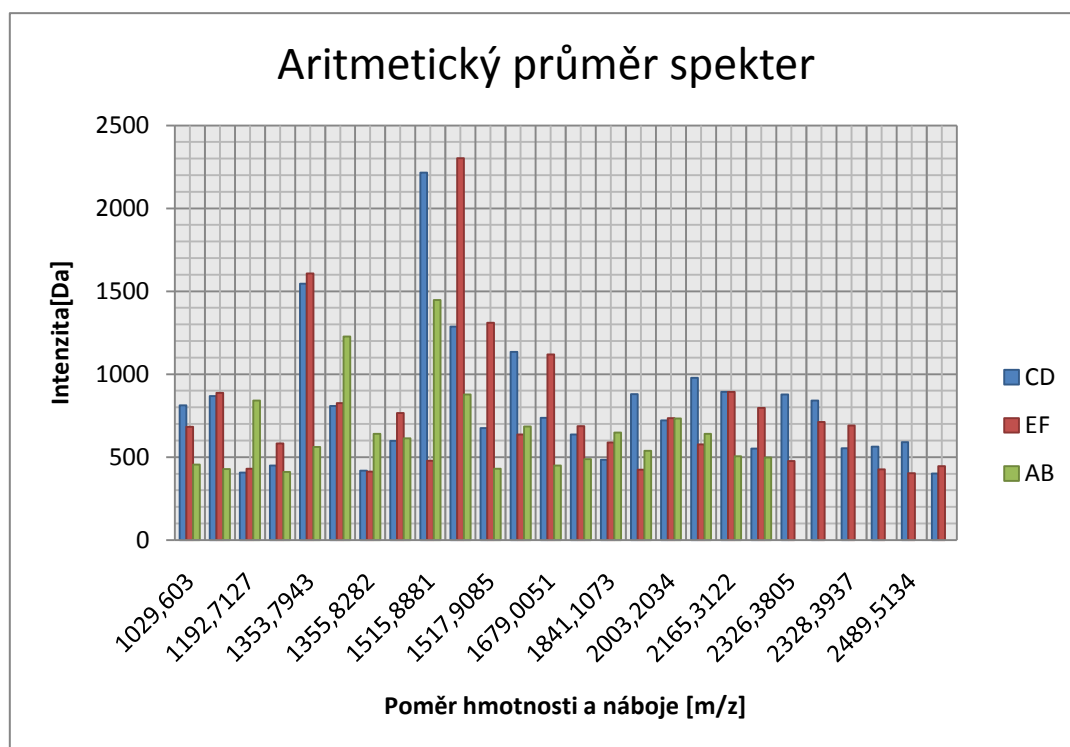
Střední hodnota		
AB	CD	EF
78,9376	101,5876	98,5178

**Tabulka 8:** Hodnoty aritmetického průměru získaných píku spekter skupiny AB, CD a EF

AB	
m/z	intenzita[Da]
1029,554	455,8628
1175,701	428,4268
1191,647	842,024
1192,659	409,9421
1337,767	561,6919
1353,736	1226,592
1354,732	639,4992
1499,842	612,9986
1515,825	1446,537
1516,836	877,276
1517,846	429,9787
1677,922	683,9882
1678,938	449,3751
1840,02	488,4288
2002,068	647,9222
2003,078	538,0781
2164,184	732,554
2165,181	639,9921
2326,297	504,5568
2327,277	497,4977

CD	
m/z	intenzita[Da]
1029,603	812,1717
1191,7	867,7695
1192,713	406,4083
1337,825	449,6866
1353,794	1545,427
1354,832	808,6627
1355,828	419,1722
1499,948	597,9008
1515,888	2215,247
1516,898	1287,963
1517,909	674,2244
1677,989	1134,722
1679,005	736,2757
1840,091	636,5917
1841,107	483,8163
2002,194	880,4327
2003,203	721,9017
2164,315	978,3196
2165,312	893,2478
2166,31	551,0529
2326,381	876,8667
2327,414	840,9357
2328,394	552,6801
2488,501	563,4375
2489,513	589,6922
2490,47	400,4585

EF	
m/z	intenzita[Da]
1029,608	683,2097
1191,706	887,6654
1192,719	429,4806
1337,873	582,4323
1353,843	1606,541
1354,839	826,4571
1355,835	413,3626
1499,955	765,549
1500,96	477,8485
1515,94	2302,778
1516,95	1309,518
1517,916	637,3159
1678,043	1119,101
1679,014	686,0781
1840,149	588,7384
1841,117	425,085
2002,255	734,6446
2003,214	576,0548
2164,326	892,3049
2165,323	796,0674
2166,321	475,4744
2326,447	711,6676
2327,426	691,4621
2328,46	426,6502
2488,513	403,4403
2489,526	445,1945



**Obrázek 34:** Graf výsledných píků spekter skupiny AB, CD a EF získaných metodou aritmetického průměru

## 8 Závěr

Bakalářská práce uvádí principy analýzy látek pomocí hmotnostní spektrometrie a tandemové hmotnostní spektrometrie s volně dostupnými nástroji pro analýzu hmotnostních spekter získaných pomocí obou metod. Dále byly popsány principy funkcí spektrometrů, jejich využití v praxi a možnosti jejich využití. V další kapitole je popsáno hmotnostní spektrum, jeho definice a interpretace. Tato část zadání je dokumentována v teoretickém úvodu práce.

V rámci spolupráce s Mendlovou univerzitou v Brně bylo v plánu asistovat při analýze vzorků flavonoidní sloučeniny pomocí hmotnostního spektrometru. Tato část zadání nebyla z technických důvodů splněna. Ze stejného důvodu nebyla získána i modelová spektra isoflavonů daidzeinu a genisteinu. Byly však získány naměřená a zpracovaná spektra z jiné databáze (spektra prokvašené a povařené mladiny a sladiny). Tato spektra byla uložena ve formátu t2d.

Aby mohla být tato spektra načtena, bylo zapotřebí využít Java knihovny v programu Proteom a vytvořit funkční propojení těchto knihoven s programem Matlab. Pro analýzu takto získaných vzorků byla vytvořena aplikace v prostředí Matlab. Následně byla pomocí této aplikace provedena matematická analýza získaných spekter.

Ze závěrů, které jsou uvedeny v kapitole 7 lze usoudit, že spektra sladiny vykazují menší obsah isoflavonů (flavonoidních sloučenin) než spektra mladiny jak povařené, tak i prokvašené. Statistické parametry popisují, že mezi zpracovanými vzorky mladiny jsou rozdíly v obsahu isoflavonů jen minimální. Vykazují ale odlišnosti v porovnání se vzorky sladiny. Grafické znázornění opět potvrzuje domněnku, že mladina je větším zdrojem isoflavonů než sladina.

Práce může být následně rozšířena. Jedná se především o načítání dalších existujících formátů souborů, které obsahují hmotnostní spektra. Dále pak hromadné načtení a zpracování všech dat v adresáři a jejich souhrnnou analýzu. Možnosti aplikace jsou prakticky neomezené, například lze vytvořit databázi hmotnostních spekter a ty pak porovnávat s aktuálně zkoumanými spektry. Statistické parametry mohou být také rozšířeny o další možnosti vyhodnocení zkoumaných spekter. Jedná se především o modus, medián, jednovýběrový t-test a další možnosti statistického vyhodnocení.

## Seznam literatury

- [1] *Mezinárodní konference o hmotnostní spektrometrii* [online]. 2006 [cit. 2009-12-19]. Dostupný z WWW: <<http://www.veda.cz/article.do?articleId=13256>>.
- [2] *New Mass Spectrometers from Thermo Fisher* [online]. 1995-2009 [cit. 2009-12-20]. Dostupný z WWW: <<http://www.labx.com/v2/adsearch/detail3.cfm?adnumb=330302>>.
- [3] *Hmotnostní spektrometrie v organické analýze* [online]. 2009 [cit. 2009-12-19]. Dostupný z WWW: <[http://holcapek.upce.cz/teaching/MS1\\_Uvod.pdf](http://holcapek.upce.cz/teaching/MS1_Uvod.pdf)>.
- [4] *Úvod do proteomiky* [online]. 2005-2006 [cit. 2009-12-19]. Dostupný z WWW: <<http://biol.lf1.cuni.cz/ucebnice/proteomika.htm>>.
- [5] ODSTRČÍLKOVÁ, H. *Anaerobní biooxidace síry. Diplomová práce* [online]. 2007 [cit. 2009-12-19]. Dostupný z WWW: <[https://is.muni.cz/auth/th/78111/prif\\_m/diplomova\\_prace.pdf](https://is.muni.cz/auth/th/78111/prif_m/diplomova_prace.pdf)>.
- [6] HERNYCHOVÁ, L. *Základy hmotnostní spektrometrie* [online]. 2005 [cit. 2009-12-19]. Dostupný z WWW: <[http://webak.upce.cz/~kbbv/Student/Vyuka/Imunologie\\_imunochemie/Prednasky/Prednasky\\_Henrychova/Zaklady\\_hmotnostni\\_spektrometrie.pdf](http://webak.upce.cz/~kbbv/Student/Vyuka/Imunologie_imunochemie/Prednasky/Prednasky_Henrychova/Zaklady_hmotnostni_spektrometrie.pdf)>.
- [7] GREGUŠ, P., NOVÁKOVÁ, L. *Vývoj diagnostické metody pro kvalitativní a kvantitativní analýzu farnesolu a tyrosolu, jako produktů metabolismu kvasinky candida albicans v biologických vzorcích s využitím techniky uplc/ms/ms* [online]. 2009 [cit. 2009-12-19]. Dostupný z WWW: <<http://kalch.upce.cz/merck09/pdf/Gregus.pdf>>.
- [8] VODIČKOVÁ, H., et al. *Analytické metody pro stanovení metalothioneinů* [online]. 2001 [cit. 2009-12-19]. Dostupný z WWW: <<http://www.chemicke-listy.cz/docs/full/archiv/2001/08-PDF/477-483.pdf>>.
- [9] VRÁBEL, P. *Spojení kapilární elektroforezy s hmotnostní spektrometrií pro analýzu peptidů. Diplomová práce* [online]. 2003 [cit. 2009-12-19]. Dostupný z WWW: <[https://is.muni.cz/auth/th/8861/prif\\_m/text.pdf](https://is.muni.cz/auth/th/8861/prif_m/text.pdf)>.
- [10] NOVÁK, J. *Aplikace metrických indexovacích metod na data získaná hmotnostní spektrometrií. Diplomová práce* [online]. 2008 [cit. 2009-12-19]. Dostupný z WWW: <[http://siret.ms.mff.cuni.cz/hoksza/studentpapers/diploma\\_jri\\_novak.pdf](http://siret.ms.mff.cuni.cz/hoksza/studentpapers/diploma_jri_novak.pdf)>.
- [11] HAVLIŠ, J. *Hmotnostní spektrometrie MALDI TOF* [online]. 1999 [cit. 2009-12-19]. Dostupný z WWW: <<http://www.vesmir.cz/clanek/hmotnostni-spektrometrie-maldi-tof>>.
- [12] *Základní principy měření molekulových hmotností* [online]. 2005 [cit. 2009-12-19]. Dostupný z WWW: <<http://biomikro.vscht.cz/maldiman/cz/theory/basics.php>>.
- [13] *Hmotnostní spektrometrie (MS)* [online]. 2005 [cit. 2009-12-19]. Dostupný z WWW: <[http://tomcat.prf.jcu.cz/sima/analyticka\\_chemie/vybranemet.htm](http://tomcat.prf.jcu.cz/sima/analyticka_chemie/vybranemet.htm)>.
- [14] VALLA, M. *Systém pro zpracování skóre z metod identifikace proteinů v tandemové hmotnostní spektrometrii. Diplomová práce* [online]. 2008 [cit. 2009-12-19]. Dostupný z WWW: <[https://www.vutbr.cz/www\\_base/zav\\_prace\\_soubor\\_verejne.php?file\\_id=8226](https://www.vutbr.cz/www_base/zav_prace_soubor_verejne.php?file_id=8226)>.
- [15] KODÍČEK, M. *Biochemické pojmy: výkladový slovník* [online]. 2005 [cit. 2009-12-19]. Dostupný z WWW: <[http://vydavatelstvi.vscht.cz/knihy/uid\\_es-002/ebook.html?p=da](http://vydavatelstvi.vscht.cz/knihy/uid_es-002/ebook.html?p=da)>.
- [16] UDRŽALOVÁ, J. *Analýza proteinů a peptidů bakterie Francisella tularensis za účelem výběru identifikačních a typizačních markerů* [online]. [s.l.], 2009. 87 s. Diplomová práce. Univerzita Pardubice. Dostupné z WWW: <[http://dspace.upce.cz/bitstream/10195/33581/1/UdrzalovaJ\\_Analyza%20proteinu\\_MH\\_2009.pdf](http://dspace.upce.cz/bitstream/10195/33581/1/UdrzalovaJ_Analyza%20proteinu_MH_2009.pdf)>.

- [17] *Matrix Scienc : Peptide Fragmentation* [online]. 2007 [cit. 2009-12-20]. Dostupný z WWW: <[http://www.matrixscience.com/help/fragmentation\\_help.html](http://www.matrixscience.com/help/fragmentation_help.html)>.
- [18] *Sekvenování bílkovin a peptidů* [online]. 2005 [cit. 2009-12-20]. Dostupný z WWW: <<http://biomikro.vscht.cz/maldivan/cz/theory/sequencing.php>>
- [19] KINTER, M., SHERMAN, N.: *Protein Sequencing and Identification Using Tandem Mass Spectrometry*. 1st edition. New York: Wiley, John & Sons, Incorporated, 2000. 301 s. New Edition. ISBN 9780471322498.
- [20] KAPP, E., SCHÜTZ, F. *Overview of Tandem Mass Spectrometry (MS/MS) Database Search Algorithms* [online]. 2007 [cit. 2009-12-20]. Dostupný z WWW: <<http://www.currentprotocols.com/protocol/ps2502>>.
- [21] DANČÍK, V., et al. *De Novo Peptide Sequencing via Tandem Mass Spectrometry* [online]. 1999 [cit. 2009-12-20]. Dostupný z WWW: <<http://www2.in.tu-clausthal.de/~hammer/biosem/paper/dancik.pdf>>.
- [22] *Mass Spectrometry and Biotechnology Resource* [online]. 2000-2009 [cit. 2009-12-20]. Dostupný z WWW: <<http://www.ionsource.com/>>.
- [23] *HUPO Proteomics Standards Initiative: mzML 1.1.0 Specification* [online]. 2007 [cit. 2009-12-20]. Dostupný z WWW: <<http://www.psidev.info/index.php?q=node/257>>.
- [24] *Seattle Proteome Center: Formats:mzXML* [online]. 2009 [cit. 2009-12-20]. Dostupný z WWW: <<http://tools.proteomecenter.org/wiki/index.php?title=Formats:mzXML>>.
- [25] *The Global Proteome Machine Organization: X!TANDEM project* [online]. 2004 [cit. 2009-12-20]. Dostupný z WWW: <<http://www.thegpm.org/tandem/index.html>>.
- [26] *Proteome Software: X!Tandem Explained* [online]. c2009 [cit. 2009-12-20]. Dostupný z WWW: <[http://www.proteomesoftware.com/Proteome\\_software\\_pro\\_xtandem.html](http://www.proteomesoftware.com/Proteome_software_pro_xtandem.html)>.
- [27] *INFOSCIENCE: Tandem Mass Spectrometry Protein Identification on a PC Grid* [online]. 2007 [cit. 2009-12-20]. Dostupný z WWW: <<http://infoscience.epfl.ch/record/101150/files/>>.
- [28] GEER, L., et al. *Reducing false positive rates in MS/MS sequence searching and incorporating intensity into match based statistics* [online]. 2006 [cit. 2009-12-20]. Dostupný z WWW: <<ftp://ftp.ncbi.nih.gov/pub/lewisg/presentations/asms06poster.pdf>>.
- [29] EER, Lewis Y., et al. *Open Mass Spectrometry Search Algorithm* [online]. 2009 [cit. 2009-12-20]. Dostupný z WWW: <<http://arxiv.org/ftp/q-bio/papers/0406/0406002.pdf>>.
- [30] Calculation of protein expectation value from peptide expectation values in X!Tandem [online]. 2004 [cit. 2009-12-27]. Dostupný z WWW: <[http://www.thegpm.org/docs/peptide\\_protein\\_expect.pdf](http://www.thegpm.org/docs/peptide_protein_expect.pdf)>.
- [31] ZOSSO, D., et al. *Infoscience* [online]. 2007 [cit. 2010-05-01]. Tandem Mass Spectrometry Protein Identification on a PC Grid. Dostupné z WWW: <[http://infoscience.epfl.ch/record/101150/files/zosso\\_healthgrid07.pdf?version=1](http://infoscience.epfl.ch/record/101150/files/zosso_healthgrid07.pdf?version=1)>.
- [32] FALKNER, J.A., et al. *Http://bioinformatics.oxfordjournals.org* [online]. 2006 [cit. 2010-05-01]. ProteomeCommons.org IO Framework: reading and writing multiple proteomics data formats. Dostupné z WWW: <<http://bioinformatics.oxfordjournals.org/cgi/reprint/btl573v1.pdf>>.
- [33] *http://proteomecommons.org* [online]. c2010 [cit. 2010-05-01]. ProteomeCommons.org IO Framework 6.21. Dostupné z WWW: <<http://proteomecommons.org/current/531/>>.
- [34] VEITCH, N. C. *Isoflavonoids of the Leguminosae*. Natural Product Reports [online]. 2007, 24, [cit. 2010-05-15]. Dostupný z WWW: <<http://www.rsc.org/delivery/ArticleLinking/DisplayArticleForFree.cfm?doi=b511238a&JournalCode=NP>>.

- [35] REYNAUD, J., et al. *Isoflavonoids in non-leguminous families: an update*. Natural Product Reports [online]. 2005, 22, [cit. 2010-05-15]. Dostupný z WWW: <http://www.rsc.org/delivery/ArticleLinking/DisplayArticleForFree.cfm?doi=b416248j&JournalCode=NP>.
- [36] PIČMANOVÁ, M. *Isoflavonoidy v nebobovitých rostlinách: fytochemie, biologické funkce a molekulární biologie* [online]. [s.l.], 2008. 36 s. Bakalářská práce. Univerzita Karlova v Praze. Dostupné z WWW: <http://kfrserver.natur.cuni.cz/studium/bakalar/prace/picmanova.pdf>.
- [37] KLEJDUS, B. *Separace a identifikace isoflavonů v rostlinném materiálu* [online]. [s.l.], 2004. 51 s. Habilitační práce. Univerzita Palackého v Olomouci. Dostupné z WWW: [http://www.upol.cz/fileadmin/user\\_upload/PrF-dokumenty/Vedicka\\_rada/Habilitace\\_a\\_profesury/ukon\\_hab\\_prof/Klejdus\\_Borivoj\\_hab/ha\\_b.prace-Klejdus.pdf](http://www.upol.cz/fileadmin/user_upload/PrF-dokumenty/Vedicka_rada/Habilitace_a_profesury/ukon_hab_prof/Klejdus_Borivoj_hab/ha_b.prace-Klejdus.pdf).
- [38] <http://www.phytochemicals.info/> [online]. 2007 [cit. 2010-05-16]. Phytochemicals. Dostupné z WWW: <http://www.phytochemicals.info/phytochemicals/genistein.php>.
- [39] <http://www.phytochemicals.info/> [online]. 2007 [cit. 2010-05-16]. Phytochemicals. Dostupné z WWW: <http://www.phytochemicals.info/phytochemicals/daidzein.php>.
- [40] JELÍNEK, J.; ZICHÁČEK, V. *Biologie pro gymnázia*. 2000. Olomouc: Nakladatelství Olomouc, 2000. 544 s. ISBN 80-7182-107-1.
- [41] <http://rostliny.prirodou.cz/> [online]. c2006-2009 [cit. 2010-05-16]. Bobovité. Dostupné z WWW: <http://rostliny.prirodou.cz/bobovite/>.
- [42] <http://www.bylinky.kvalitne.cz/> [online]. 2010 [cit. 2010-05-16]. Bylinky. Dostupné z WWW: [http://www.bylinky.kvalitne.cz/fotky/jetel\\_lucni.jpg](http://www.bylinky.kvalitne.cz/fotky/jetel_lucni.jpg).
- [43] ARNDT, T. *Červený jetel a fytoestrogeny* [online]. 2008 [cit. 2010-05-16]. Dostupné z WWW: <http://www.celostnimedicina.cz/cervený-jetel-a-fytoestrogeny.htm>.
- [44] MIKELOVÁ, R., et al. *Chromatografické stanovení isoflavonů ve vegetativních a generativních částech rostlin sóje (glycine max)*. CHEMagazín [online]. 2004, XIV, 1, [cit. 2010-05-16]. Dostupný z WWW: <http://user.mendelu.cz/~kizek/publikace/pdf/mikelova.pdf>.
- [45] <http://popinave-rostliny.atlasrostlin.cz/> [online]. c2010 [cit. 2010-05-16]. Dostupné z WWW: <http://popinave-rostliny.atlasrostlin.cz/chmel-otacivy>.
- [46] ZVÁROVÁ, J. *Biomedicínská statistika: I. Základy statistiky pro biomedicínské obory*. 1. vydání. Praha: Univerzita Karlova v Praze - Nakladatelství Karolinum, 2007. 215 s. ISBN 978-80-7184-786-1.
- [47] NOVOVIČOVÁ, J. *Pravděpodobnost a matematická statistika* [online]. [s.l.] : Vydavatelství ČVUT, 1999 [cit. 2010-05-18]. Dostupné z WWW: <http://euler.fd.cvut.cz/publikace/files/skripta3.pdf>.
- [48] The MathWorks [online]. 1984-2010 [cit. 2010-05-21]. *Bioinformatics Toolbox*. Dostupné z WWW: <http://www.mathworks.com/access/helpdesk/help/toolbox/bioinfo/ref/>
- [49] FABRIK, I. *Kombinace separačních metod s hmotnostní spektrometrií pro charakterizaci sacharidů* [online]. [s.l.], 2010. 77 s. Diplomová práce. Masarykova univerzita. Dostupné z WWW: [https://is.muni.cz/auth/th/175253/prif\\_m/DP\\_Fabrik\\_2010.pdf](https://is.muni.cz/auth/th/175253/prif_m/DP_Fabrik_2010.pdf).

## Seznam obrázků

Obrázek 1: Hmotnostní spektrometr, THERMO FISHER LTQ Orbitrap XL .....	5
Obrázek 2: Základní schéma hmotnostního spektrometru .....	6
Obrázek 3: Princip metody ESI.....	7
Obrázek 4: Schéma metody MALDI .....	8
Obrázek 5: Hmotnostní spektrum .....	10
Obrázek 6: Tandemová hmotnostní spektrometrie .....	12
Obrázek 7: Určení C- a N- konců peptidu .....	13
Obrázek 8: Klasifikace iontů.....	13
Obrázek 9: Immoniový iont .....	14
Obrázek 10: Schéma sekvenování peptidů spektrometrem MALDI-TOF.....	15
Obrázek 11: PSD MALDI-TOF spektrum peptidu.....	15
Obrázek 12: Strukturní charakterizace hmotnosti vybraného iontu pomocí MS/MS .....	16
Obrázek 13: Pracovní tok algoritmu .....	19
Obrázek 14: Srovnání hmotnostních spekter .....	19
Obrázek 15: Výběr hyper-skóre .....	20
Obrázek 16: Určení hranice statistické významnosti H .....	21
Obrázek 17: Grafické určení E-value.....	22
Obrázek 18: Vývojový diagram algoritmu .....	25
Obrázek 19: Histogram spekter.....	27
Obrázek 20: Histogramy falešně pozitivních a pravdivě pozitivních výsledků.....	29
Obrázek 21: Základní struktura flavonoidů .....	31
Obrázek 22: Difenylypropanový skelet (a), základní skelet flavonu (b) a isoflavonu (c) .....	32
Obrázek 23: Základní struktura isoflavonů.....	33
Obrázek 24: Genistein.....	33
Obrázek 25: Daidzein.....	33
Obrázek 26: Jetel luční ( <i>Trifolium pratense</i> ) .....	34
Obrázek 27: Sója luštinatá ( <i>Glycine max</i> ).....	35
Obrázek 28: Chmel ( <i>Humulus lupulus</i> ) .....	37
Obrázek 29: Zjednodušené schéma výroby piva.....	38
Obrázek 30: Zobrazení píků pomocí funkce Mspeaks.....	41
Obrázek 31: Zobrazení píků pomocí funkce Msheatmap .....	41
Obrázek 32: Zobrazení csv dat v programu MS Excel .....	43



Obrázek 33: Vzhled GUI aplikace .....	43
Obrázek 34: Graf výsledných píků skupiny AB, CD a EF získaných metodou aritmetického průměru .....	47

## Seznam tabulek

Tabulka 1: Přehled dostupných implementací pro metodu "De Novo" .....	16
Tabulka 2: Přehled dostupných implementací pro metodu "Sequence Tag" .....	17
Tabulka 3: Množství isoflavonů v různých částech rostliny sóji luštinaté ( <i>Glycine max</i> ) .....	35
Tabulka 4: Přehled funkcí Bioinformatického toolboxu pro zpracování MS spekter.....	40
Tabulka 5: Korelační koeficient spekter skupin AB, CD, EF .....	45
Tabulka 6: Rozptyl spekter skupin AB, CD, EF .....	45
Tabulka 7: Střední hodnota spekter skupin AB, CD, EF .....	45
Tabulka 8: Hodnoty aritmetického průměru získaných píků spekter skupiny AB, CD a EF .....	46

## Seznam příloh

<i>Příloha 1</i>	Základní vlastnosti aplikací využívající pro identifikaci peptidů neinterpretovaná tandemová spektra
<i>Příloha 2</i>	Přehled struktury isoflavonů
<i>Příloha 3</i>	Vývojový diagram aplikace vytvořené v grafickém uživatelském rozhraní (GUI) programu Matlab.

## Příloha 1

Základní vlastnosti aplikací využívající pro identifikaci peptidů neinterpretovaná tandemová spektra [20]

Název	Dostupnost	Internetová adresa	Online přístup <sup>a</sup>
SEQUEST	Komerční	<a href="http://fields.scripps.edu.sequest">http://fields.scripps.edu.sequest</a>	Ne
MASCOT	Komerční	<a href="http://www.matrixscience.com">http://www.matrixscience.com</a>	Ano
SPECTRUM MILL	Komerční	<a href="http://www.chem.agilent.com">http://www.chem.agilent.com</a>	Ne
PLGS	Komerční	<a href="http://www.waters.com">http://www.waters.com</a>	Ne
PEAKS	Komerční	<a href="http://www.bioinformaticssolutions.com">http://www.bioinformaticssolutions.com</a>	Ano
PROTEINPILOT	Komerční	<a href="http://download.appliedbiosystems.com/proteinpilot">http://download.appliedbiosystems.com/proteinpilot</a>	Ne
PHENYX	Komerční	<a href="http://www.phenylx-ms.com">http://www.phenylx-ms.com</a>	Ano
SORCERER	Komerční	<a href="http://www.sagensearch.com">http://www.sagensearch.com</a>	Ne
OMSSA	Open-source, volně šiřitelný	<a href="http://pubchem.ncbi.nlm.nih.gov/omssa/">http://pubchem.ncbi.nlm.nih.gov/omssa/</a>	Ano
X!TANDEM	Open-source, volně šiřitelný	<a href="http://www.thegpm.org/TANDEM">http://www.thegpm.org/TANDEM</a>	Ne

<sup>a</sup> Online přístup znamená, že vyhledávání může být uskutečněno vzdáleně a přistupovat lze prostřednictvím webového prohlížeče.

## Příloha 2

Přehled struktury isoflavonů [37]

Název isoflavonů	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	R <sub>4</sub>	R <sub>5</sub>	R <sub>6</sub>
Daidzin	H	H	glc	H	H	OH
glycetin-7-O-β-D-glucosid	H	OCH <sub>3</sub>	glc	H	H	OH
kalycosin-7-O-β-D-glucosid	H	H	glc	H	OH	OCH <sub>3</sub>
genistin	OH	H	glc	H	H	OH
daidzien-7-O-β-D-glucosid-6''-O-malonát	H	H	glc-Mal	H	H	OH
3-methylorobol-7-O-β-D-glucosid	OH	H	glc	H	OCH <sub>3</sub>	OH
pratensein-7-O-β-D-glucosid	OH	H	glc	H	OH	OCH <sub>3</sub>
kalykosin-7-O-β-D-glucosid-6''-O-malonát	H	H	glc-Mal	H	OH	OCH <sub>3</sub>
pseudobaptigenin-7-O-β-D-glucosid	H	H	glc	H	O-	OCH <sub>2</sub> -
daidzein-7-O-β-D-glucosid-6''-O-acetát	H	H	glc-OAc	H	H	OH
ononin (formononetin-7-O-β-D-glucosid)	H	H	glc	H	H	OCH <sub>3</sub>
genistein-7-O-β-D-glucosid-6''-O-malonát	OH	H	glc-Mal	H	H	OH
orobol-7-O-β-D-glucosid-6''-O-malonát	OH	H	glc	H	OH	OH
3-methylorobol-7-O-β-D-glucosid-6''-O-malonát	OH	H	glc-Mal	H	OCH <sub>3</sub>	OH
pratensein-7-O-β-D-glucosid-6''-O-malonát	OH	H	glc-Mal	H	OH	OCH <sub>3</sub>
daidzein	H	H	H	H	H	OH
irilone-4'-O-β-D-glucosid	OH	O-	CH <sub>2</sub> -	H	H	O-glc
pseudobaptigenin-7-O-β-D-glucosid-6''-O-malonát	H	H	glc-Mal	H	O-	OCH <sub>2</sub> -
glycitein	H	OCH <sub>3</sub>	H	H	H	OH
orobol	OH	H	H	H	OH	OH
kalykosin	H	H	H	H	OH	OCH <sub>3</sub>
formononetin-7-O-β-D-glucosid-6''-O-malonát	H	H	glc-Mal	H	H	OCH <sub>3</sub>
afroformosin-7-O-β-D-glucosid	H	OCH <sub>3</sub>	glc	H	H	OCH <sub>3</sub>
sissostrin (biochanin A-7-O-β-D-glucosid )	OH	H	glc	H	H	OCH <sub>3</sub>
irilin B-7-O-β-D-glucosid	OH	OCH <sub>3</sub>	glc	OH	H	H
irilone-4'-O-β-D-glucosid-6''-O-malonát	OH	O-	CH <sub>2</sub> -	H	H	O-glc-Mal
trifoside (prunetin-4'-O-β-D-glucosid)	OH	H	CH <sub>3</sub>	H	H	O-glc
afroformosin-7-O-β-D-glucosid-6''-O-malonát	H	OCH <sub>3</sub>	glc-Mal	H	H	OCH <sub>3</sub>
pseudobaptigenin-7-O-β-D-glucosid-6''-O-acetát	H	H	glc-OAc	H	O-	OCH <sub>2</sub> -
formononetin-7-O-β-D-glucosid-6''-O-acetát	H	H	glc-OAc	H	H	OCH <sub>3</sub>
texasin -7-O-β-D-glucosid-6''-O-malonát	H	OH	glc-Mal	H	H	OCH <sub>3</sub>
irilin B-7-O-β-D-glucosid-6''-O-malonát	OH	OCH <sub>3</sub>	glc-Mal	OH	H	H
3'-methylorobol	OH	H	H	H	OCH <sub>3</sub>	OH
genistein	OH	H	H	H	H	OH
biochanin A-7-O-β-D-glucosid-6''-O-malonát	OH	H	glc-Mal	H	H	OCH <sub>3</sub>
pratensein	OH	H	OH	H	OH	OCH <sub>3</sub>
prunetin-4'-O-β-D-glucosid-6''-O-malonát	OH	H	CH <sub>3</sub>	H	H	O-glc-Mal
pseudobaptigenin	H	H	H	H	O-	OCH <sub>2</sub> -
irilone-4'-O-β-D-glucosid-6''-O-acetát	OH	O-	CH <sub>2</sub> -	H	H	O-glc-OAc
formononetin	H	H	H	H	H	OCH <sub>3</sub>
prunetin-4'-O-β-D-glucosid-6''-O-acetát	OH	H	CH <sub>3</sub>	H	H	O-glc-OAc
texasin	H	OH	H	H	H	OCH <sub>3</sub>
biochanin A-7-O-β-D-glucosid-6''-O-acetát	OH	H	glc-OAc	H	H	OCH <sub>3</sub>
irilone	OH	O-	CH <sub>2</sub> -	H	H	OH
prunetin	OH	H	CH <sub>3</sub>	H	H	OH
biochanin A	OH	H	H	H	H	OCH <sub>3</sub>

## Příloha 3

Vývojový diagram aplikace vytvořené v grafickém uživatelském rozhraní (GUI) programu Matlab.

